

What strategy for European science?

The European Commission has grown to be a powerful influence on research, and should not now be surprised that its activities will in future be scrutinized more closely. Winning the respect of the research community is its chief task.

THE European Commission, the executive arm of the European Union (EU), used to be a mere bit player in research. When its direct spending amounted to less than 2 per cent of spending in its member states, the frequent waywardness of its decisions on research could be overlooked. But the fourth 'framework' programme takes spending on research to 4 per cent of the total, while the common practice of 50:50 funding of research projects means that the commission probably calls the shots for 6 per cent of all Europe's spending on research. The figures show that national contributions to European collaborations not under the EU's wing (of which the European Space Agency is the most expensive) account for a further 7 per cent of all research spending. This is no longer an inconspicuous proportion. Moreover, the commission's pocket is deep enough to allow it to influence other pan-European institutions, the European Molecular Biology Laboratory (EMBL) for example. (The commission may be instrumental in keeping Italy in EMBL if it agrees to support the planned mouse-genetics laboratory near Rome — see *Nature* 372, 305; 1994.)

In these changed circumstances, the European research community, and Europe more widely, needs to know that the commission is spending its money wisely and well. But there are several structural impediments to the good management of research from Brussels. One is that the commission has no coherent and comprehensive policy on research, and even lacks a means of formulating and maintaining one. To be sure, Professor Antonio Ruberti, the commissioner for research for the past two years, has put flesh on the bones of the old plan to create a committee of 100 of Europe's great and good in research, called the European Science and Technology Assembly (ESTA), and would no doubt be willing to listen to what it has to say. But Ruberti will be replaced in January by Mme Edith Cresson, who may have different ideas and a different agenda for research support.

Continuity

Continuity is not the commission's strong suit, although chunks of the research budget administered by other than the commission's research directorate appear to acquire an uncanny permanence. Thus the commission's Esprit programme of collaborative research in information technology, which dates back to 1980, has survived intact the Single Act of 1986 (which makes Europe into a true common market), the passage of 15 years and the general suspicion that it would not survive hard-headed evaluation. The con-

tinuation of Esprit probably owes more to horse-trading among the separate parts of the commission than it does to common sense.

Policy

But times change. When the Maastricht Treaty was ratified last year, the European Parliament acquired rights of scrutiny (and veto) over the research budget. Strictly speaking, the parliament and the Council of Ministers (the EU's policy-making body, representing national governments) are partners in the determination of policy, but they are still tussling between themselves about their mutual balance of power. That will further complicate the determination of policy and budgets in the years ahead. There will be horse-trading between the council and the parliament as well as within the commission. In those circumstances, the opinions of the research community in Europe about the directions that policy will take are likely to be further marginalized.

That prospect partly explains the formation of the body with the ungainly name of Eurohorcs, the 'European heads of research councils' (see *Nature* 372, 306; 1994), which represents the organizations responsible for national governments' support for research, and which can be counted on to scrutinize the commission's doings, and even to tell their parent governments when they disapprove of what they find. The commission will thus find that winning approval for future framework programmes will be more difficult than in the past. Overlapping membership will probably ensure that ESTA sings the same tune. But, if so, the tune will probably again be that of organizations, not of people at the bench.

That tendency will be further reinforced by the moves now under way to enlarge the commission's responsibility for research in the EU. The Maastricht Treaty ratified at the end of last year gives the commission responsibility for 'coordinating' research in Europe, whereas its previous responsibility was simply for the encouragement of cooperation between member states and their research organizations. The working paper on the subject, entitled "Achieving coordination through cooperation" (and dated 19 October 1994), shows that the commission is conscious of the delicacy of its task. To begin with, it will ask member states merely for information, and will attempt to use its own network (and especially the new European Science and Technology Observatory at Seville in Spain) to provide them, in turn, with useful information. Further ahead lies the prospect of extra funding for particular coordinated projects.



That is a sensible approach, but again one in which the opinions of the research community are likely to be represented by those of national government agencies. The two are not necessarily identical. Moreover, in the drive for coordination, there will inevitably be a risk that the value of research projects will be equated simply with the value, to industry or more broadly, of the knowledge they generate. But, as the whole world knows, research projects (especially in an academic setting) are the most effective way of generating the technical skill for which the modern world is crying out. Over-effective coordination could be the death of that.

That is one reason why the research community will be rightly suspicious of the commission's plans. Another is more practical: the commission has a poor track record in persuading the research community of the good sense of its decisions. That has much to do with the administrative muddle that often attends the award of research grants, but also with the sense that officials in Brussels have undue influence on the grants actually awarded even when these have been through a serious bout of peer-review (which process is not always serious). The practical goal, should be to conduct business in such a way as to persuade even those grant-seekers whom the commission must disappoint that their case has been considered fairly.

That, of course, is housekeeping and not strategy. What, beyond that, should the commission aim to do? One obvious need is to strengthen research in member states where it is at present neglected (Greece, Ireland and Portugal, for example). Doing that equitably will not be easy. Another is the pursuit of research in which the whole of Europe has an interest. With the emergence of bovine spongiform encephalopathy (BSE) as a disease of British cattle, and continued speculation about the risks to beef-eating people, should not the commission take an interest in prion diseases? Indeed, a competence in public health more generally would also be worthwhile. There is also a need for a mechanism for providing components of the research infrastructure intermediate in cost between projects such as those at CERN (the European Laboratory for Particle Physics) and those that national governments can conveniently provide. The station now established at Dwingeloo in the Netherlands for extracting meaningful data from very-long-baseline interferometric observations is only one of many examples.

In all these activities and others, the commission will succeed only if it can persuade the whole of Europe's research community of the good sense of its spending plans. But that will be more and not less difficult in the years immediately ahead. With the power struggle between the parliament and the council certain to intensify, the commission will be tempted to skimp on its more public role of carrying the community with it. That, after all, is how it has earned the reputation of being a *dirigiste* entity in other fields. The trouble, in research, is that the research community is likely to be even more resentful of what seems like top-down direction than are, say, European farmers. The commission would be wise to go carefully. Mme Cresson should, above all, be seen to be listening not just to ESTA but also to all the other bodies ready to offer her advice. □

Pitfalls of co-authorship

A disputed article in a British journal again reveals that co-authorship is not a free gift.

THE British evidently have some way to go before they become sensitized to the improper use of the scientific literature. So much is clear from the circumstances surrounding the publication last August of a paper in the *British Journal of Obstetrics and Gynaecology* (101, 716–717; 1994) by J. M. Pearce, I. Manyonda and G. J. Chamberlain. The paper, in the journal's idiom a "case-report", described the transfer of an ectopic pregnancy to the uterus of the woman in whom it had developed; the paper claimed that the outcome was the successful birth of a child. Evidently the result would have been of some importance.

Doubts have been raised about the accuracy of the report on largely administrative grounds, notably that nobody at St George's Hospital in London can recall having assisted in the operation. The hospital has mounted an inquiry, as a consequence of which the first author of the paper, Mr Malcolm Pearce, has been suspended pending a decision on "what further action will be taken" by the hospital and the medical school associated with it. Pearce denies impropriety, and must be taken at his word while the findings of the hospital's inquiry remain secret (which is also proper while the hospital has not decided what further action, if any, to take).

But there is one issue that, even at this stage, will cause consternation in the research community. The third author is otherwise Sir Geoffrey Chamberlain, and happens to be both the current president of the Royal College of Obstetrics and Gynaecology and editor-in-chief of its journal, the *British Journal of Obstetrics and Gynaecology*. Pearce, on the other hand, is a member of Chamberlain's staff and one of several deputy editors of the same journal, which is published on behalf of the Royal College by Blackwell Scientific Press. Neither Chamberlain nor the third author, Dr I. Manyonda, is mentioned by the hospital as having been the subject of the inquiry carried out.

Indeed, Pearce is reported to have said that Chamberlain's name was included among the authors of the paper "purely as a courtesy, as head of the department". If true, that is an extraordinary statement, entirely at odds with the evidence that has accumulated elsewhere that 'honorary co-authorship' is a disreputable and also a dangerous practice. The fraudulent publications more than a decade ago attributable to John C. Darsee, first at Emory University and then at the Harvard Medical School, so sullied reputations at both institutions that honorary co-authorship is now widely regarded as a potentially poisoned chalice. It is also a misdemeanour in its own right, attributing credit where it does not belong. In the unfortunate case at St George's, it is important that the hospital should eventually say whether Chamberlain knew he had been cited as a co-author. The Royal College (which has set up an "independent" inquiry) should urgently take steps to rid its journal of insider publishing. □

ORI finds Imanishi-Kari guilty of misconduct, proposes 10-year ban

Washington. A long-awaited report released in Washington this week finds Theresa Imanishi-Kari, co-author with the Nobel laureate David Baltimore of an controversial immunology paper, to be guilty of 19 charges of scientific misconduct. It recommends that she be barred from receiving federal research funds for 10 years.

The report was issued by the Office of Research Integrity (ORI), part of the Department of Health and Human Services. It is a culmination of nine years of extraordinary investigations which have involved the Congress, forensic experts from the Secret Service, and investigators from the National Institutes of Health's former Office of Scientific Integrity.

The report, which includes about 200 pages of forensic evidence, statistical analysis and scientific discussion, took three and a half years to complete. This time span has itself been criticized. "It is an excellent report," says one congressional source. "But it is unconscionable that it took so long."

But even now the matter is not yet closed. Imanishi-Kari has filed an appeal, and has told Tufts University in Massachusetts, where she now works, that the ORI's findings are "totally unfounded".

David Baltimore, professor of molecular biology and immunology at the Massachusetts Institute of Technology — who re-

signed as president of Rockefeller University in New York following criticism from colleagues of his defence of Imanishi-Kari — told *Nature* on Monday that "to my knowledge, she never did anything wrong". He added: "I believe the ORI report is totally wrong in accusing her of misconduct."

But Lyle Bivens, director of the ORI, said after a moment's reflection that she found Baltimore's comment "baffling", adding: "I'll stand by the report."

The specific conclusions of the report are that Imanishi-Kari fabricated and falsified critical parts of her published results, and that she compounded this by denying the misconduct and fabricating data that "she claimed supported her initial findings". The report also concludes that she "further compounded the fabrications and falsifications by referencing or reporting them in grant applications submitted to the National Institutes of Health".

Imanishi-Kari's research group had been investigating how a gene from a B-cell of one strain of mice would affect the type of antibodies produced by the immune system of another strain. The work constitutes one approach to understanding the regulation of the immune system, and could in the long run have implications for understanding how to confer resistance to disease.

The controversy started in 1986 when

Margot O'Toole, a post-doctoral research fellow in Imanishi-Kari's laboratory at MIT, noticed that data in a laboratory notebook did not support those that had been published. She claims that she was also put under pressure to report her own data in a misleading fashion. "From the beginning I just wanted a correction," says O'Toole. "I didn't want anyone debarred."

O'Toole discussed the discrepancies with senior scientists and subsequently with officials at MIT and Tufts. Both universities reviewed the paper, MIT because Imanishi-Kari worked there, and Tufts because she was a candidate for a post at the university.

MIT found no evidence of deliberate falsification or fabrication, and Tufts agreed that O'Toole had identified an error but that it was not flagrant. (The ORI is now undertaking an investigation into how the two universities handled O'Toole's allegations).

Shortly afterwards, the National Institutes of Health put together a panel to investigate the affair. This found "significant errors of misstatement or omission", but concluded that there was no evidence of fraud or conscious manipulation.

At about the same time, however, the House subcommittee on oversight and investigation, chaired by John Dingell (Democrat-Michigan), subpoenaed the original data, and opened its own investigation. In addition, according to the ORI report, the NIH discovered that Imanishi-Kari's notebooks "had not been compiled contemporaneously with the conduct of the reported experiment".

The NIH reopened its investigation, and its Office of Scientific Integrity completed a draft report in March 1991, which found scientific misconduct based on falsification and fabrication. The US Attorney in Maryland considered a criminal prosecution. When no prosecution resulted, the newly formed ORI, which had replaced the Office of Scientific Integrity, took up the case. Its findings are similar to, but more extensive than, those of the OSI.

O'Toole, who for a long time was cold-shouldered in academic circles, is now a research scientist working at the biotechnology company GI in Boston. "It is so hard to bring evidence of scientific misconduct to light," she says. "Even if there is incontrovertible evidence, there are monumental hurdles."

O'Toole is one of four whistleblowers that the Commission on Research Integrity has asked to testify at hearings being held this week. "My main point will be about the importance of the integrity of data," she says.

Helen Gavaghan

UK 'protects' health research funds

London. Virginia Bottomley, Britain's health secretary, last week announced that the government had agreed to allocate an additional £40 million (\$62 million) to university medical schools to help cover the costs of research in teaching hospitals.

Bottomley's decision is in line with her earlier endorsement of proposals made in a report prepared by a team under Anthony Culyer, professor of medical economics at the University of York, that health department funding for research should be channelled into a single, identifiable stream.

In making the announcement, Bottomley said that the additional money being provided directly to the medical schools will allow work to be speeded up within the National Health Service's

research and development programme in "innovative and evidence-based health care."

But health department officials admit that the extra support for research does not represent 'new money'. An equivalent amount will be taken away from that which is provided to local authorities and other health 'purchasers' for buying the services of teaching hospitals, which will be expected to reduce the prices that they charge to health care purchasers accordingly.

While acknowledging the government's plan that this should lead to a more level playing field for the providers of health care — i.e. by bringing the costs charged by teaching and non-teaching hospitals closer in line — many medical schools fear they could lose out.

"The extra £40 million is a welcome contribution," says Michael Powell of the Committee of Vice-Chancellors and Principals, the body which represents university medical schools in their negotiations with government. "But universities would like to see it go much further than that." □

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

**Bottomley: money
will be ring-fenced.**

Energy labs face uncertain future after US elections

San Francisco. Researchers at California's two largest government-funded research laboratories fear that they may face major budget cuts — or possibly even elimination — under the Republican-dominated Congress elected last month.

A major part of the platform on which the Republicans won control of Congress was a promise to balance the federal budget, cut taxes and slash government spending. According to both Republican and Democrat sources in Washington, this could result in considerable pressure on the Lawrence Berkeley and Lawrence Livermore laboratories, despite the new majority party's commitment to basic science.

Supporters of the laboratories fear that even if the Republicans keep their promise to restore military spending, this will do little for research, because most of the money will go directly to improving the readiness of the armed forces.

Some Republican officials have already indicated that they would like to break up the \$18.5-billion Department of Energy (DoE), which manages both laboratories through the University of California. Representative John Kasich (Republican, Ohio), although a supporter of defence spending, has in the past proposed closing the federal laboratories on the grounds that they are no longer needed. He is expected to become chair of the powerful House of Representatives Budget Committee.

The DoE, which was created in 1977 by former President Jimmy Carter and took over the laboratories of the former Atomic Energy Commission, has come under attack before. But it may not survive this time.

Diminishing concern about the need to separate weapons research from the military — initially intended to enhance the security surrounding research and development on nuclear weapons — means that it would be easier to move such work to the Department of Defense, and to divide the DoE's other programmes among other agencies. Such a change would have a dramatic effect on Lawrence Berkeley and Lawrence Livermore, which rely on the energy department for 80 per cent of their funding.

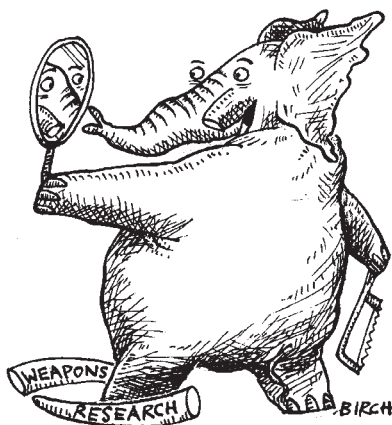
Nuclear weapons research — which accounts for about half of Lawrence Livermore's \$868-million operating budget — would probably move to the defence department, with the rest split between agencies. The laboratory might then find itself with less powerful political support against further budget cutting than in the past.

Other programmes likely to be closely scrutinized are those aimed at transferring technology to industry. Under the Clinton administration, these have grown in impor-

tance at both laboratories, as a means of keeping research teams intact and boosting the US economy.

But many Republicans view such programmes as government interference in the market place. The Advanced Technology Program, organized by the National Institute of Standards and Technology (NIST), has been targeted for particular attack.

Research programmes on renewable energy, energy efficiency and climate change could be cut back severely, although the



human genome programme is likely to be spared. With industry's help, the laboratories hope to persuade the new Congress that technology transfer is not the same as industrial subsidy, and deserves continued support.

The severity of the cuts will become apparent during the first half of next year as Congress shapes the federal budget for the year starting 1st October 1995, on the basis of a budget proposal due from President Bill Clinton in February.

A report on the management of the weapons laboratories, commissioned last year by Energy Secretary Hazel O'Leary from a committee chaired by Robert Galvin, chairman of the electronics company Motorola, is due for publication in February, and is also likely to affect the budget outcome (see *Nature* 371, 368; 1994). Although the task force preparing it has not revealed its thinking, a source close to the laboratories says he feels the group was impressed by their research, but worried about the quality of the DoE supervision of the laboratories.

Additional uncertainty over the future of the laboratories arises from the new and unknown priorities of those tipped to chair the congressional committees responsible for their funding. But some of the core defence missions of Lawrence Livermore and the basic research work at Lawrence Berkeley could still find support among Republicans.

Sally Lehrman

Italian drug firms rethink ban on conference support

Munich. Italian pharmaceutical companies look likely to reverse a temporary ban on the commercial sponsorship of scientific congresses following complaints from the scientific community.

In Italy, about a third of the funding for scientific conferences comes from industrial sponsors, and scientists fear that some meetings planned for next year may have to be cancelled, including the first European Congress of Pharmacology, scheduled to be held in Milan next June.

The ban was introduced as part of the industry's response to health reforms introduced by the government in January. The reforms were aimed in particular at reducing the prices of key drugs, and at limiting the number of drugs whose purchase is reimbursed by the state.

Following a drastic fall in profits, Farindustria, the umbrella association representing the Italian pharmaceutical industry, decided in October on a six-month ban on all promotional activities. If the financial situation facing the industry does not improve, warns Farindustria spokesman Cesare Fassari, the ban may be extended.

Drug companies have agreed to stop advertising in medical journals and to cut the distribution of free drug samples to doctors by 60 per cent. They have also decided to stop sponsoring scientific congresses — which such companies often use as a way of promoting their products — for the first half of 1995.

By supporting a uniform ban across the industry, individual drug companies hope that they will be able to save money without losing their competitive edge in the Italian market. They also hope to regain public confidence in the industry and thus discourage any further cuts in drug prices, after corruption scandals last year (see *Nature* 364, 658; 1993).

But scientists are upset that genuine scientific conferences, which have nothing to do with product promotion, are included in the ban. After a meeting with scientists several weeks ago, Farindustria is now considering reversing the ban. A small advisory committee representing scientific societies will assess congresses seeking sponsorship, and will make recommendations to Farindustria to fund those with genuine scientific content.

Gianni Marini, vice-president of Farindustria, confirms that the board of Farindustria will soon decide whether to make exceptions for meetings with a genuine scientific content which have already been planned. "I am in favour," he says. "We cannot interfere with scientific communication."

Toni Feder

German minister backs national academy

Munich. Jürgen Rüttgers, Germany's new minister of education, science, research and technology, promised in his inaugural speech to the Bundestag, the federal parliament, last week that the government would increase its funding of research and technology.

In a speech that won the backing of the main opposition party, the Social Democrats (SPD), he also promised to promote the government's efforts to create a national academy of sciences, to reduce the burden of regulation on the country's scientists and to reduce the amount of time spent by German students at university before entering employment.

But achieving such goals may be easier said than done — existing regional bodies will oppose the new academy and shortly after Rüttgers' speech, the federal finance minister indicated that there would be no extra money for science in 1995.

Rüttgers was speaking the day after the German chancellor, Helmut Kohl, had outlined the government's plan for the next four years. Kohl said that one of its chief goals was to make the country "fit for the twenty-first century", and added that it needed "an alliance for the future" — seen by some as an allusion to the informal name of Rüttger's new department as the 'ministry for the future'. Rüttgers said that while maintaining Germany's generous funding of basic science, he planned to spend more on the development of key technologies. Another top priority, he said, was to tackle overcrowding in universities.

Patents stalemate on biotechnology

Paris. Attempts to reach agreement on a Europe-wide directive aimed at harmonizing legislation on biotechnology patents among the 12 member states of the European Union reached an official stalemate on Monday, when representatives from the Council of Ministers and the European Parliament failed to agree to a single text.

Agreement through such a 'conciliation committee' is required under the new procedures of the Maastricht treaty. But the Council of Ministers, which represents the governments of the EU states, refused to accept an amendment from the parliament that would have forbidden patents on isolated parts of the human body — including genes and cell lines. The parliament refused to budge either, reflecting its deeply held views on this issue (see *Nature* 372, 310; 1994).

If a compromise is to be reached, it will have to be found quickly. Procedures provide only weeks in which to hold another 'conciliation' meeting. If this committee eventually rejects the parliament's amendment, its decision could still be overturned by an absolute majority vote in the parliament, but many may not wish the issue to be taken to a vote. **Declan Butler**

Referring to Germany's notoriously over-bureaucratized regulatory laws on issues such as the use of chemicals and research on animals, Rüttgers said that science "should be given the breathing space it needs", promising to reduce barriers to creativity and competitiveness.

He also acknowledged that there were limits that scientific progress should respect — particularly those relating to the dignity of the individual.

Rüttgers has long been a proponent of plans to establish a national academy of science, modelled on that of the United States and intended to provide a both a national focus for the scientific community, and a point of contact with similar organizations in other countries. He was also largely responsible for setting up the

Bundestag's office of technology assessment (see *Nature* 372, 311; 1994).

In his speech to parliament, Rüttgers said he would like the proposed academy, which would represent both natural and social sciences, to serve as a forum for discussing scientific problems at a national and international level. But learned societies in many of Germany's *Länder* have given this idea a cool reception, fearing that a national academy would detract from their own prestige and challenge their independence.

The SPD's support for Rüttgers' initiatives is likely to be particularly important in the case of university reforms, which reflect a long-standing goal of Chancellor Kohl. Universities are the responsibility of the *Länder*, and the majority of *Länder* prime ministers are social democrats. □

Protests greet Italian budget cuts

Rome. A cross-party initiative in the Italian parliament is expected to block an attempt by the research minister, Stefano Podestà, to impose stringent cuts on the budget of the National Agency for New Technology, Energy and the Environment (ENEA).

The proposed cuts are the only seriously contentious issue in proposals put forward by Podestà last month for a no-growth budget for research in Italy for 1995. The budgets of the National Research Council (CNR) and the Italian Space Agency (ASI) are likely to be held at the same level as this year, namely L1,040 billion (US\$655 million) and L850 billion respectively.

Only the National Institute for Nuclear Physics (INFN) is earmarked for a rise, from L400 billion to L470 billion. But Luciano Maiani, head of the institute, says this is primarily intended to compensate for a 10 per cent cut this year.

Furthermore, the INFN will have to divert five per cent of its 1995 budget to a fund administered by the research ministry to support research in small and medium-sized enterprises for work related to physics.

But ENEA, which Podestà considers inefficient and lacking in focus, was singled out for a cut of 21 per cent, from L508 billion to L400 billion. Initially established to develop nuclear energy,

the agency was forced to change direction when Italy voted in a referendum 10 years ago to abandon further development of nuclear power.

Nicola Cabbibo, ENEA's president, acknowledges that the change in research direction was not easy for the large agency, which has in the past employed up to 5,000 staff. He also says that although ENEA has been left with a broad portfolio of activities, this does not mean that it has no focus.

Since his appointment last July, Cabbibo has been working to improve efficiency and reduce staff numbers, and he says that he is upset by Podestà's efforts to try to force his hand by simply cutting off support.

Cabbibo claims that his views have substantial support in parliament and is optimistic that his agency will be given a reprieve.

In parallel to the budget proposal, Podestà has also put forward a bill to create an interministerial committee to restructure Italian research. No details about how this committee might work have been published. But the bill could give the government stronger powers to create or close institutes without consulting the research organizations themselves.

It would also probably force external evaluations on all research institutes that are not subject to such independent scrutiny at present, a plan introduced by Podestà's predecessor, Umberto Colombo, earlier this year.

Cabbibo would welcome the idea of external evaluations of ENEA institutes if members of evaluation committees were from abroad. He fears that because the move is tightly linked to budget discussions, the motivation for the evaluations is to make further cuts. But he is reserving final judgement until more details are known.

Alison Abbott



Maiani: "Increase for physics is illusory."

Agreement near on new telescope in Namibia

Cape Town. Representatives from South Africa, Namibia and Germany are meeting this week in Windhoek to discuss plans to erect a 3.5–4.0-metre telescope on the Gamsberg, a flat-topped mountain 150 km southwest of the Namibian capital.

This has emerged as the most likely site for the Southern African Large Telescope (SALT), which has been on the drawing board for the past six years. The European Southern Observatory (ESO), based at Garching near Munich, is also interested in the site: it is the most favoured alternative site for the Very Large Telescope if land claims prevent its erection at Paranal in Chile (see *Nature* 370, 494; 1994). ESO officials have been conducting tests on atmospheric turbulence at Gamsberg since July.

The Namibian government has already granted outline approval for the project. The task of the meeting will therefore be to draft agreements between the three governments and the two institutions involved: the South African Foundation for Research Development (FRD) and the German Max Planck Institute for Astronomy, which is based in Heidelberg.

One day of the meeting will be spent inspecting the site itself. This is a plateau approximately 2 km square and 2,350 metres above sea level, which the Max Planck Institute bought in the early 1970s from a local farmer.

South African scientists would prefer the telescope to be erected at the FRD's South African Astronomical Observatory (SAAO)

at Sutherland, about 400 km northeast of Cape Town. But requests to the South African cabinet for funding over the past six years have not been approved.

"If the Germans are prepared to fund the telescope, and they have a preference for the Gamsberg, then we are happy with that", says Bob Stobie, director of SAAO and one of the South African representatives at the Windhoek meeting. Stobie adds that "Gamsberg is also a better site, but only marginally so".

A feasibility study by the SAAO has estimated the cost of the telescope at R95 million (US\$27 million). An additional R13 million would be required for developing the necessary infrastructure at the Gamsberg site (such infrastructure is already in place at Sutherland).

The SAAO was built jointly by Britain and South Africa in 1972, when its opening was attended by Lady (Margaret) Thatcher, then the UK secretary of state for education and science. But Britain suspended the joint programme in 1986 and has said that it is not prepared to contribute to the cost of a new telescope.

Under the terms of the new agreement between the FRD and the UK Particle Physics and Astronomy Research Council, a substantial amount of the British government's equipment that remained at Sutherland will now formally be handed over to South Africa.

Turbulence is the critical factor that can prevent angular resolution measured from the ground from approaching that which can be measured from space.

The main justification for the construction of the new telescope is that many objects, including the Large and Small Magellanic Clouds and the Milky Way, can be seen more easily from southern latitudes.

Longitude is also crucial; certain events may be observable only from the African continent, as other major southern land masses may be in daylight when they occur. In addition, observations from Africa can provide continuous coverage of certain astronomical phenomena.

The angular resolution of the new telescope will be an order of magnitude better than the existing one. This is only 1.9 metres long and is almost 50 years old, having first been erected at the Radcliffe Observatory in Pretoria in 1948, before being moved to Sutherland.

But policies on the allocation of time at Sutherland have allowed astronomers to make longer-scale observations than at most observatories, and this has contributed greatly to its success. Stobie feels that if used in a similar manner, the new 3.5–4-metre telescope will even be able to compete with the new generation of 8-metre telescopes.

Michael Cherry

Tokyo University puts levy on gifts

Tokyo. Hiroyuki Yoshikawa, the president of Tokyo University, has won the approval of the university's senate to set up a special 'presidential fund' based on a one per cent levy on donations that faculty members receive each year from private industry.

The fund will amount to only ¥60 million (US\$600,000) a year, and therefore represents a relatively modest step towards achieving greater financial autonomy both for Yoshikawa's office and for his university, which is the leading national university in Japan. But the move could have repercussions throughout Japan's university system. At present, Japan's national universities are strictly regulated and financed by the Ministry of Education, Science and Culture. As a result, they have very limited general funds at their own disposal.

Donations from private industry provide a substantial source of funds for Japan's universities, amounting to more than ¥50 billion a year. This is only slightly less than the government's annual budget for research grants distributed to universities by the ministry of education.

The money is given by industry chiefly to help it recruit good students from relevant departments. The donations grew rapidly during the bubble economy of the late 1980s but growth has slowed during the recession (see *Nature* 360, 98; 1992).

But the money is distributed in tiny parcels averaging about ¥1 million to indi-

vidual faculty members, and is not pooled for the university as a whole. For example, the ¥6.17 billion donated to Tokyo University last year was made up of 5,973 separate donations.

Unlike grants provided by the ministry of education, which can be used only for very narrowly defined purposes, researchers have considerable freedom in using their donations. One of the most common uses is for attending conferences and overseas travel, an area that receives poor support from the government.

By pooling a percentage of the donations, Yoshikawa will create a central fund free from the bureaucratic strings of the government. As such it is being welcomed as a dramatic step forward by reformists in universities and the government — and seems likely to be imitated by other universities.

But Yoshikawa's proposal has still faced strong opposition within the ranks of his own university. In particular, members of the faculty of engineering, of which Yoshikawa was dean before taking over the presidency, as well as those of the faculties of science and medicine, have opposed the idea.

Their opposition is not surprising, as these faculties receive by far the largest share of donations. In contrast, the proposal was warmly welcomed by other faculties, such as education and arts, which receive very few donations from industry.

For the present, the use of the fund will be limited by the senate to paying for overseas trips by faculty members undertaking study tours related to the formation of new institutes or organizations within the university. But Yoshikawa is hoping to use it for more general purposes in the future.

David Swinbanks



Yoshikawa: setting a precedent in Japan

Science foundation ponders its post-Maastricht role

Strasbourg. The European Science Foundation (ESF), Europe's 'bottom-up' research organization whose members are national research funding agencies, is still seeking a new role, although its annual assembly in Strasbourg last week failed, despite its efforts, to achieve significant progress towards this goal.

A year ago, ESF had provisionally decided that it would seek the role of the voice of European science at Brussels. But that prize has been snatched from its grasp by the creation of the European Science and Technology Assembly (ESTA).

ESF has been further hindered from achieving its goal by the formal creation of Eurohorcs — the European Heads of Research Councils — which plans to monitor and influence the European Commission's research programmes. That blow has been particularly hurtful, as the members of Eurohorcs are among the ESF's own member organizations.

Other organizations are also competing for influence in this field. The Academia Europea, with more than 1,500 elected members is too powerful for ESF to ignore, according to Lord Flowers, ESF's first president when it was founded 20 years ago. And the newly-formed Alliance of European Academies (ALLEA) is also knocking on the door of Brussels.

But Sir Dai Rees, president of ESF, told the foundation it should not lapse into "graceful retirement". He described the task of the ESF as being to help its member organizations collaborate with each other on international projects.

ESF has a good track record in this field. For example, having held its first European research conferences five years ago, the foundation now functions as a subcontractor to the EU's Training and Mobility programme in the organization of conferences. This year, there will have been 38; next year's total should amount to 50.

Networking is another ESF activity that is flourishing, based on the principle of creating a network of European researchers around a chosen topic. Dervilla Donnelly, the vice-president of the foundation and professor of chemistry at University College, Dublin, boasts that the networks often evolve into research projects that qualify for continuing support from the EU.

Some also recruit funds from outside Europe, such as the taxonomy network, which has US members and support. In future, networks will be required to produce a final document to suggest what should happen next.

The more substantial research programmes, whose cost to ESF ranges from FF100,000 to FF1.6 million (US\$18,000 to US\$300,000) a year, again involve Europe-wide collaboration and cover a wide spectrum. ESF's committees have now been asked to identify fields in which future research projects can be mounted profitably.

Last week's assembly also amended the ESF's statutes to remove the possibility that countries such as those of Central and Eastern Europe might acquire associate status, and heard the auditors rap the management over the knuckles for making too much use of overnight courier services. **John Maddox**

Transport ban raises new hurdle to nuclear waste store

Paris. A local court in northern Germany last week blocked the transport of the first consignment of nuclear waste to the controversial interim storage repository at Gorleben from the Phillipsburg power station in Baden-Württemberg. The decision buys time for Germany's anti-nuclear movement, which is using Gorleben as a hostage in its efforts to prevent the continued use of nuclear power in the country.

The interim storage repository has stood empty for 11 years. It forms part of a huge complex at Gorleben that also includes a pilot conditioning plant for compacting waste for burial in the adjacent planned deep storage repository.

The storage of nuclear waste is not itself the main issue; even if Germany closed all its nuclear power stations tomorrow, existing waste — and that from dismantled plants — would still have to go somewhere.

But blocking waste storage is one element in a two-pronged strategy by anti-nuclear groups to persuade Germany to pull out of nuclear power altogether. The other has been to stall the opening of plants for producing mixed oxide fuel (MOX), on the grounds that industry uses reprocessing to reduce waste, but needs MOX to dispose of the plutonium this produces. MOX plants would therefore also commit Germany to continue using nuclear power.

"We don't want to separate the question of waste from that of the future of nuclear power," says a spokesperson for Greenpeace. So far the strategy has been effective; the political uncertainty it has created has led to a *de facto* moratorium on the building of new power stations in Germany.

Last week a court in Lüneberg, Lower Saxony, ruled that no waste could be brought to Gorleben pending the outcome of a challenge by local residents to the interim repository's permission to operate. But the company that runs the site, Brennelement-lager Gorleben, has appealed to a higher court, which will rule in January.

Uta Kreutzenbeck, a spokeswoman for the ministry of environment of Lower Saxony — which opposes nuclear power — says that the state is also concerned about the safety of the 'Castor' flasks used to transport and hold the waste. It refused a previous attempt to deliver nuclear waste on these grounds in July, but was later overruled by the pro-nuclear federal government.

Kreutzenbeck also argues that the feasibility of safe long-term storage of nuclear waste has not been demonstrated anywhere in the world, and that there is therefore "no sense" in interim storage. "Waste should remain at the power stations," she says.

Declan Butler

EC commissioner leaves his mark

Strasbourg. Antonio Ruberti, the European Union's research commissioner who will return to Italy as a professor of systems analysis at the University of La Verna at the end of the year, is doing everything he can to leave a clear desk for Mme Edith Cresson, the former French prime minister, who is due to succeed him in January.



Ruberti: heading back to La Verna

Ruberti is proud of his two-year stint as a member of the European Commission. A likeable man, he has persuaded most people to like him. In Strasbourg to address the European Science

Foundation, Ruberti says that the European Science and Technology Assembly (which he set up) is going to be "an independent voice" within the European Union.

But his chief remaining task is to win the approval of the Council of Ministers of a policy to coordinate national research policies within the EU, in order fully to implement the Maastricht Treaty. "Articles 139 H, K & L" repeat all those in Ruberti's circle who know the reference.

Ruberti is confident that he will leave a commission more aware of the importance of research than when he joined, and remains passionate on the importance of links between research and the wider culture. He pleaded last week for an open system of science and technology that would, among other things, be responsive to the economic needs of Europe. **J. M.**

Court asked to halt inquiry into psychologist's suicide

Montreal. Quebec's Superior Court has been asked to suspend an inquiry ordered by McGill University into the death of the neuropsychologist Justine Sergent, who committed suicide last April (see *Nature* 369, 176; 1993). It has also been asked to remove the lawyer that the university appointed to carry out its inquiry.

Both requests have come from Michael Pesner, the accountant charged with settling Sergent's estate, whose remit includes protecting Sergent's reputation. He claims that Caspar Bloom, the president of the Montreal bar who is conducting the inquiry, should not do so because he is being paid by the university, and his son is a resident physician at a McGill teaching hospital.

The university refuses to comment on the petition until a judge hears a request for a temporary injunction, due to take place this week. In addition to Bloom's inquiry into McGill's handling of the Sergent case, a scientific audit of her work is under way and is expected to be completed by the end of the year.

Sergent and her husband Yves committed suicide after publication of a newspaper article triggered by an anonymous letter that revealed that she had been reprimanded by the university for failing to observe regulations on the use of human subjects in experiments. She was contesting the reprimand.

As part of Pesner's petition, Sergent's suicide note was published for the first time. In it she gave her version of the dispute with the ethics committee of the Montreal Neurological Institute. "I have been accused of breaking rules," she wrote. "The ethics committee of the Montreal Neurological Institute [MNI] considered that I did not abide by the rules because I carried out a study on reading musical notation with positron emission tomography [PET]."

She says that, as the ethics committee had approved experiments using PET technology in subjects with faces, letters or words

as stimuli, she "did not consider putting subjects at more risk by exposing them to a different type of visual information".

Her note also claimed that the ethics committee was not competent to decide whether or not she broke a rule because it did not follow guidelines laid down by the Medical Research Council of Canada. These, she said, called for at least one expert in her field to sit on the committee evaluating her research, but she claimed that none did so. (A spokeswoman for the university said she was unable to verify whether such an expert had participated in the committee's work.)

Sergent also admitted in her note to altering the telephone numbers of some subjects. But she said that she did this because the consent forms signed by the subjects stipulate that everything about participation in a study is confidential.

"I had received complaints from subjects participating in earlier studies that someone, claiming to be either from 'security at McGill' or from the 'director's office' was calling them and asking them questions about the various tasks they had been performing during my PET studies," she said.

"I therefore decided to ensure the confidentiality of the subjects by altering the telephone number on the official form which I did in their presence, while keeping the correct number in my files."

Publication of the newspaper story had discredited both her work and her career in a way "which I cannot tolerate", she said. Before the publication of the article, she had hoped that the audit of her research would clear her name, as the letter had raised doubts "about my honesty as a scientist".

But she said she could not be assured that she would be treated fairly by the MNI, claiming that the institute's director had withheld from an evaluation committee information she had sent to him in connection with a request to be considered for a promotion.

David Spurgeon

Highway in the sky comes to the aid of India's scientists

New Delhi. Indian scientists, frequently frustrated by the high cost of travel and subscriptions to foreign journals, are turning enthusiastically to a high-speed, satellite-based information highway with which they can communicate instantly and free of charge between themselves and counterparts on INTERNET in 150 countries.

Commissioned by the National Informatics Centre (NIC), the data-link now connects 15 cities. NIC's director-general, N. Seshagiri, says it will link a total of 70 cities and towns — serving more than 10,000 users — by the end of 1996.

Voice, texts, graphics and photographs can be transmitted over the network at 2.2 million bits per second (b.p.s.), a speed that Seshagiri claims conforms to the standard international definition of an 'information highway'.

Seshagiri says that scientists can use the highway to send and receive electronic mail and news of conferences and seminars, to transmit research papers for publication, and to hold video conferences. Bibliographic information on 15,000 journals will eventually be available, while scientists will also be able to obtain full texts of articles from 1,000 journals, including graphics.

Scientists from 2,000 selected academic and research institutions have already been invited to use this facility. By the middle of 1996, the network's services will be available "at the doorsteps of all of India's 8,000 research, educational and medical institutions", says Seshagiri.

The overall cost of the highway will be less than £10 million. Seshagiri says this was achieved by avoiding expensive fibre optics, by using NIC's existing satellite-based network NICNET as a backbone, and by limiting its applications. "We have designed our highway for scientists and businessmen — not to provide video on demand or for shopping from home," he says.

NICNET was set up in 1988 and links government computers in 500 cities and towns through the INSAT domestic satellite, using nearly 700 roof-top antennas. These operate in the C-band (5 MHz), sending messages at 9,600 b.p.s., a speed that is too slow for multimedia applications.

Seshagiri says that part of NICNET, linking important centres, has been turned into an information highway by installing high-speed satellite terminals operating at 36 MHz. As INSAT does not carry any transponders operating in the relevant wave band, the highway relies on an Intelsat satellite. But NIC will switch to the domestically produced INSAT-2C — with an appropriate transponder — when it is launched in 1995.

K. S. Jayaraman

Salk Institute president is replaced

Washington. Brain Henderson has resigned as president and chief executive of the Salk Institute, the biomedical research centre, based in San Diego, California founded by polio vaccine pioneer Jonas Salk.

Henderson left the Institute suddenly in October, returning full-time to his cancer epidemiology research at the University of Southern California. He denies reports that he was fired: "The challenge of fund-raising for the institute and doing science at the same time was tougher than

I thought," he says. "I realised I could no longer do both."

The trustees of the prestigious institute moved quickly to split Henderson's function, and last week named Charles Massey as new chief executive officer and Francis Crick — winner of the Nobel Prize with James Watson for discovering the DNA double helix — as president.

The Salk Institute, which specialises in molecular biology, genetics and neuroscience, has fifty senior faculty and eight hundred support staff. Colin Macilwain

France plans national gene therapy strategy

Paris. France's ministries of health and of research have announced their intention to coordinate their support for genome research and its clinical applications under the umbrella of a single so-called 'Génome-Santé' programme.

The initial goals of the joint programme, which will also involve medical charities such as the influential French Muscular Dystrophy Association (AFM), will include strengthening fundamental research in gene therapy, creating gene therapy centres at major hospitals, and increasing the emphasis on genetics in medical training.

The new integrated approach has been welcomed by both medical charities and researchers in gene therapy. "It's a radical change of government attitude," says Bernard Barataud, the president of the AFM, which has been a driving force in French genome efforts, and which created the Généthron laboratory in 1990 with the Paris-based Centre d'Etudes du Polymorphisme Humain (CEPH). Barataud, who has long criticized genome efforts in France as being too fragmented, says there is now the beginning of "a vision of the whole".

The most immediate measure is a call for research proposals in gene therapy, to be funded jointly by the two ministries, the National Agency for AIDS Research, AFM and the National League Against Cancer. This will provide around FF35 million (US\$6.6 million) in 1995, and aim, in particular, to support the creation of a handful of gene therapy centres associated with major hospitals.

Priorities will include fundamental research on animal models of genetic diseases and on new vectors. This shows that the government accepts that progress depends on easing bottlenecks in fundamental research, says Alain Fisher, who leads France's largest gene therapy group at the Hôpital Necker in Paris.

Fisher, who was a co-author of a recent report to the government on gene therapy, also says he is pleased that the joint call for proposals will be peer reviewed by a committee that includes foreign scientists.

Under the plans, France's 43 medical schools would also create new staff positions for teaching new genetics courses. Only 15 schools have well developed genetics programmes, says François Fillon, the science minister, who also intends to create a new specialized diploma in genome studies.

As expected, the government has also decided to simplify regulatory procedures for gene therapy. In particular, a single interministerial commission will take over the work of the existing labyrinth of regulatory committees (see *Nature* 371, 550; 1994).

Another committee will be created to regulate DNA and gene banks, and oversee the contractual arrangements for their eventual commercial exploitation. This follows the public row earlier this year over a proposal that would have given Millennium, a Massachusetts-based biotechnology company, access to a DNA bank at CEPH in Paris (see *Nature* 370, 4; 1994).

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Généthron's gene library will be key element.

But enthusiasm for the government's proposals for gene therapy has been tempered by wider concerns. One is that, although the government is already setting up a national committee to 'coordinate' life sciences research among the various research bodies, with genome research a

declared priority, budgetary pressure may undermine France's commitment to maintaining a strong research base in life sciences in the research organizations.

Some scientists point out, for example, that the life sciences budget (excluding salaries) of the Centre National de la Recherche Scientifique (CNRS), France's principal research organization, has fallen by more than 14 per cent in less than two years. Funding for the biomedical research organization INSERM is also decreasing.

Furthermore, concern that research organizations may lose control over scientific choices to those more interested in short-term results is being stimulated by the growing influence of medical charities, whose resources now exceed the FF900 million a year that the government spends on genome research.

Fisher claims there is a potential risk that the state will abandon its responsibilities to support long-term research, and that the presence of charities may lead to too much emphasis on applied research aimed at short-term goals. Moreover, although some charities — including AFM — operate rigorous peer-review systems, he is concerned that others may distribute their money unwisely.

Declan Butler

UK 'should abandon quarantine law'

London. The United Kingdom's strict quarantine laws should be immediately replaced with a system of vaccination and blood-testing for pet cats and dogs, according to a report published last week by the Agriculture Committee of the House of Commons.

The committee says that scientific advances made since the last review of the laws in 1971, such as the development of safe vaccines and accurate serological tests, mean that it is "not only feasible but desirable" for the United Kingdom to adopt a scheme similar to those already operated in mainland Europe, and introduced by Norway and Sweden in May.

But the recommendations received a cool reception from William Waldegrave, the agriculture minister. He said he would look carefully at the recommendations, but he emphasized that alternatives would have to be shown to offer at least as much protection as the present system.

Veterinary groups were equally cautious, arguing that replacing quarantine rules immediately would be premature. "It's the cost and the anxiety and the worry of living with rabies that we want to avoid," says Neal King, president of the Royal College of Veterinary Surgeons.

Under the recommendations, six months' quarantine would still apply to cats or dogs coming from non-approved countries. But

others would avoid quarantine if they had been vaccinated with an approved inactivated rabies vaccine, had lived in an approved country for at least six months, were identified by tattooing or micro-chipping, and had a blood test showing rabies immunity four months after vaccination.

But the Royal College of Veterinary Surgeons and the British Veterinary Association (BVA) argue that Britain lacks the necessary infrastructure to police such a system adequately. Every point of entry into the United Kingdom would have to have a special animal channel, as well as facilities to carry out an animal health check.

They are also worried about adequate identification and certification: tattoos can be tampered with or the same number can be used on different animals, and microchip systems would have to be compatible for all approved countries. Paul DeVile, president of the BVA, also says that he is worried about the authenticity of certification.

"Certainly the advances over the last few years have been quite remarkable, both with the efficacy of vaccines and as regards the oral vaccination of foxes," says DeVile. But he added that he would like to see rabies eliminated from the European Union before quarantine restrictions were relaxed.

Maggie Verrall

Forensic DNA typing dispute

SIR — Lander and Budowle¹, in declaring the end of the controversy over the forensic application of DNA technology, have presented a piece of propaganda that completely distorts the current situation in a very difficult matter at the nexus of science and law.

The errors of the article begin with its title. DNA patterns are not fingerprints, because unlike fingerprints, they are not idiotypes. The struggle to resolve the issues is not a “war”, in which all “weapons” are fair, but for some of us, at least, an attempt to find out what is true. Most important, the controversy is most certainly not laid to rest but in some ways has become more urgent because of technical changes in the methods.

There are three outstanding issues. First, there is a very serious problem of laboratory reliability of DNA technology. Lander and Budowle dismiss this problem, assuring us that laboratory practice is now being scrupulously overseen by, of all agencies, the Federal Bureau of Investigation (FBI). It is the FBI laboratory, from which Dr Budowle comes, that failed to replicate its own DNA profiles in 12 per cent of samples, in an internal test of its procedures², and which has consistently refused to allow independent third-party quality control of its work. It is ludicrous to set this fox to guard the henhouse.

The problem of laboratory reliability has been greatly exacerbated by the increasing use of PCR technology to amplify small samples. Typically, a large fresh sample of blood from the accused is worked on in the same laboratory or by the same technician as a minuscule scraping of blood or tissue from the crime scene. But, as everyone who uses the PCR technique knows, the probability of a false match by contamination of the minuscule crime scene sample, to be amplified, from the large sample taken from the accused, by mislabelling, aerosols, carelessly unchanged pipette tips and other similar laboratory sloppiness, is very great and many orders of magnitude greater than the tiny match probabilities calculated for forensic purposes. Moreover, the polymerase chain reaction technology is incapable of detecting new alleles that may be possessed by the suspect or present in the crime scene sample, but which have not yet been included in the test array. In the absence of blind proficiency testing, quality-control protocols and unannounced periodic checks designed and supervised by disinterested parties, the test results from federal, state and local crime laboratories, and private contract laboratories must be regarded as unreliable.

The second problem, that of correctly calculating probabilities when there is population heterogeneity, is dealt with by

Hartl in the accompanying letter.

Finally, because DNA profiles are not idiotypes, juries are asked to consider a numerical probability of matching. But there is an extensive literature about the ability of lay people to understand probability statements and to make decisions that accord with the logic of such statements (see, for example, ref. 3). The clear result of these investigations is that people not trained in quantitative methods cannot understand issues of statistical independence and the basic logic of probability statements. For example, it is common for people to believe that a 1 in 4 chance means that the event is bound to happen on the fourth trial. Nor can a one-time instruction by a judge be sufficient to correct these misunderstandings. Because juries are no more capable of interpreting probability statements than they are of interpreting any other piece of highly technical information, there are insuperable barriers to their use in the courts.

Given the DNA polymorphism in humans, it is within our reach to design sequencing methods that provide idiotypes, uniquely identifying individuals as real fingerprints do. Coupled with quality-control checks and proficiency testing by disinterested monitors, a system of forensic identification could be created that would on the one hand protect the innocent and, on the other, help to convict the guilty. So why does the FBI not devote its time and energy to such developments, instead of trying to defend and shore up a basically flawed system? Why did Lander and Budowle choose to embrace in the pages of a leading journal of science, just before Budowle is scheduled to appear before tens of millions on television as a witness for the prosecution in what is surely the most publicized crime since the assassination of John Kennedy? As the French say, it gives one to think.

R. C. Lewontin

*Museum of Comparative Zoology,
Harvard University, Cambridge,
Massachusetts 02138, USA*

SIR — It is welcome news that the FBI, in an astonishing 180° change of direction, has agreed to adopt the interim ceiling principle recommended by the National

Research Council (NRC)¹.

The reversal of policy is long overdue in view of the acceptance of the approach by many other DNA-typing agencies. The ‘interim’ ceiling principle, based on racial databases, is a stop-gap measure intended to be replaced by a more refined method based on databases from diverse ethnic groups. Alas, Lander and Budowle’s announcement ending the DNA ‘war’ contains no indication that the FBI intends to implement the refined form of the ceiling principle. Furthermore, numerous not-so-subtle hints suggest that even the interim ceiling principle may soon be abandoned.

Ask yourself why an article about the adoption of the interim ceiling principle would attack the biological rationale for the method? The ceiling principle was designed to compensate for the genetic substructure that exists in human populations. But Lander and Budowle assert that the evidence for substructure in the Lewontin-Hartl paper⁴ is “flawed” (see ref. 5), and that the FBI’s population surveys have found only “modest” differences among ethnic groups. How do these statements square with Budowle’s previous admission⁶ that it is “universally accepted that substructure exists within major population groups”? Indeed, statistically significant differences among ethnic groups have been documented for markers used in DNA typing (see, for example, ref. 7), including many comparisons among ethnic groups in the FBI’s own population surveys.

Statistical significance is an objective, unambiguous, universally accepted standard of scientific proof. When differences in allele frequencies among ethnic groups are statistically significant, it means that they are real — the hypothesis that genetic differences among ethnic groups are negligible cannot be supported. On what basis, then, does Budowle ignore substructure? Because the differences between ethnic groups are not “forensically significant” (p. 533 of ref. 6). When does a statistically significant difference become “forensically significant”? Well, “when the likelihood of occurrences of the DNA profile would be meaningfully different”⁸. And guess who decides whether differences are “meaningfully different” . . . ?

A short life for the ceiling principle is also intimated by the statement that the controversy over population substructure would have been disposed of several years ago had the FBI been given authority to set up a committee to rule on the issue. It would have “made short work of the population genetics issue, by clarifying, changing or discarding the original NRC recommendations” (ref. 1). In fact, the FBI has been granted such authority in the DNA Identification Act of 1994. Will “short work” be made of the interim ceiling principle?

1. Lander, E. & Budowle, B. *Nature* **371**, 735–738 (1994).
2. US vs Yee, 12 F.3d 540 (6th Circuit), 1993.
3. Palmerini, M. *The Illusion of Knowing* (Wiley, Chichester, 1993).
4. Lewontin, R. C. & Hartl, D. L. *Science* **254**, 1745–1750 (1991).
5. Hartl, D. L. & Lewontin, R. C. *Science* **260**, 473–474 (1993).
6. Budowle, B., Monson, K. & Giusti, A. M. *Am. J. hum. Genet.* **55**, 533–539.
7. Krane, D. E. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10583–10587 (1992).
8. Budowle, B. & Monson, K. in *DNA Fingerprinting: The State of the Science* (eds Pena, S. D. et al.) 171–191 (Birkhauser, Basel, 1993).

There are still other indications that are troubling. Why else insist that the FBI's product-rule calculation yields the "best estimate" instead of calling it what it is — an estimate more nearly at the nonconservative end of the spectrum that can be used in conjunction with the ceiling principle to bracket the true value? Why else attempt repeatedly to marginalize the issue of population substructure by labeling it as "purely academic"? Why else brush aside many other important issues, such as proficiency testing, accessibility of databases, and technical considerations relating to polymarkers and use of the polymerase chain reaction?

Daniel L. Hartl

Department of Organismic
and Evolutionary Biology,
Harvard University,
Cambridge, Massachusetts 02178, USA

Fighting fire with fire

SIR — Colin Macilwain¹ recorded the following memorable remark of a US politician: "we pay agencies to fight fires, not light them". In Western Australia, forest managers² have for three decades been using prescribed fires to minimize the possibility of destructive forest fires similar to those that recently affected the western United States.

The bushfire hazard in the southwest of Western Australia is arguably potentially more severe than in any other region of the world. This is because of its unique tall forests, which shed tonnes of highly flammable material each year, and a Mediterranean climate, that is, a climate that produces each year a 3–6 month drought with periods of high temperature and low humidity.

Fire is a naturally occurring element of southwest ecosystems, as natural as rain and sunlight. To the vegetation and native animals, it is simply a powerful disturbance factor from which, in time, the natural systems recover, or in the presence of which they evolve. Fire can also be an agent of death and destruction to human assets and values, and large wild fires are extremely costly.

In the southwest forests alone there have been more than 200 lightning-caused fires since 1988, any of which, without the benefit of prior planning burning, might have erupted into wildfires of the scale that devastated forests in Idaho. Fuel (fallen leaves, twigs, bark and sticks) is the only factor that can be managed, and this is Western Australia's first method of defeating wildfire. The other is a fast response by well-trained firefighters. The first method, reducing the fire's energy source, makes the second effective.

Prescribed, low-to-moderate intensity fires are not ecologically harmful³ — such fires were lit every year by aboriginal

Western Australians until shortly after settlement by Europeans in 1829. Indeed, millennia of frequent firing of Western Australian forests by aborigines resulted in one of the most productive and valuable hardwood forests in the world.

Syd Shea

Department of Conservation
and Land Management,
Australia II Drive, Crawley, 6009,
Western Australia

1. *Nature* **370**, 585 (1994).
2. Shea, S. R. et al. in *Fire and the Australian Biota* (eds Gill, A. M., Groves, R. H. & Noble, I. R.) 443–470 (Aust. Academy of Science, 1981).
3. Christensen, P. & Abbott, I. *Aust. For.* **52**, 103–121 (1989).

Russian bomb

SIR — I wish to clarify your report (*Nature* **370**, 85; 1994) about the open letter by ten Russian scientists, of whom I was one.

Your report says that those responsible for building the atomic bomb in the Soviet Union were motivated by the fear that, if they failed, they would be dismissed as idealists and suffer the same fate as biologists had done under Josef Stalin. The more explicit text of our letter says that "under the conditions that existed during the totalitarian regime of Stalin, a failure of the first test of the Soviet atomic bomb would certainly have led to harsh punishment of all of Kurchatov's team and to the total devastation of Soviet physics (as happened shortly before in 1948 to Soviet biology)". We went on to say that the devastation would have begun with the prohibition of relativity theory and quantum mechanics on the grounds that they were the "poisoned tools of ideological saboteurs" and the atomic test in August 1949 "in fact saved Soviet physics".

Moreover, the main spirit of our letter was dictated not by history but by the present situation in our country: we consider Sudoplatov's book and its wide coverage as typical examples of activities of "reactionary, antidemocratic, anti-intellectual forces" that use the attacks on science "not only to discredit our intelligentsia but also to create political tensions internally and internationally, to produce an atmosphere of xenophobia and to generate hostility and mistrust within the world scientific community".

Vitaly Goldansky

Institute of Chemical Physics RAS,
4 ul. Kosygina,
Moscow 117977, Russia

News and reality

SIR — Your leading article "How to deal with dreadfulness" (*Nature* **370**, 313–314; 1994) suggests that what concerns you is what appears "dreadful" to those who follow the news media, rather than the

experience of those of us who live every day with the events that interest the reporters. You mention the "wishes of the television-watching world" as if they interest you more than the wishes of those of us who live the events.

I have lived through many of these events for 14 years and notice a great disparity between the world in which I live and the world that is reported. The small subset of the actual events that is reported is selected by the reporters and by those skilful at guiding them. There are many distortions. Photographs get miscaptioned. Lenses can make a rifleman look as if he is aiming at someone at whom he is not aiming. Interviews, including voice recordings, are edited selectively. And even if one is quoted directly, one's success or failure at conveying one's experiences accurately is limited by one's ability to choose words quickly and appropriately, and by the ability of the audience to imagine a world they have never experienced. An expert at communications will sway the "television-watching world" better than a simple honest person who does not know how to express himself.

Governments, and the United Nations, make decisions aimed at pleasing the television watchers, and based largely on the data that appear in the news media. (Reports from "observers" may be discounted. When I soldiered in Lebanon I noticed that the United Nations (UN) troops never dared leave their base. And UN observers in Hebron see what they are shown.)

So I oppose your conclusion that the UN ought to intervene in the affairs of member states, undermining "to some degree" their sovereignty. Rather than intervening militarily in far-off events about which we can know little with any degree of scientific accuracy, we should each of us try to solve problems closer to home.

Frank J. Leavitt

Kiriat Arba 303/2,
Hebron, Israel

Help wanted

SIR — The Committee on Medical Aspects of Food Policy of the Department of Health has set up a working group to prepare guidelines on the nutritional assessment of infant formulas. The group will be chaired by Dr Peter Aggett, head of nutrition, diet and health at the Institute of Food Research in Norwich. The working group invites concise written submissions based on reasoned argument and scientific data from interested parties. Submissions should be sent to me.

P. Clarke

Department of Health,
Room 593D Skipton House,
80 London Road, London SE1 6LW, UK

A scientific agenda for climate policy?

Sonja A. Boehmer-Christiansen

The complementary interests of climate scientists, national and international bureaucracies and politicians have so far determined the political dynamics of the global warming debate. But cracks are now beginning to appear.

CONTROVERSY continues to surround the Framework Convention on Climate Change, signed at the Earth Summit in Rio in 1992. While preparations are under way for the first meeting of the signatories to the convention, due to be held in Berlin next year, the scientific community, represented by the Intergovernmental Panel on Climate Change (IPCC), is already under fire for its delay in coming up with satisfactory advice.

Greater scientific certainty over climate change is unlikely until early in the next century. Indeed, there are doubts over whether we shall ever know enough about climate change in advance of the policy decisions needed to head off potential dangers. But policymakers continue to hope that, with sufficient funding, the appropriate scientific knowledge can be produced according to a timetable.

The climate treaty requires industrial countries to try to stabilize their national greenhouse gas emissions at 1990 levels by the year 2000. No binding targets beyond 2000 have been agreed; indeed, hopes are fading that this can be achieved globally. At Rio, the treaty was left deliberately imprecise to ensure both that the United States signed, and that the entire issue remain open to future research results.

Reductions in emissions have already been achieved, though primarily as a result of the recession and (for example in the former East Germany) deindustrialization. In some countries, the rapid replacement of coal and oil by less carbon-intensive fuels may be sufficient to achieve stabilization of emissions. But energy systems are difficult to turn around, and both development goals and sociopolitical expectations are slow to change.

The problem lies not with nature but with society. Given this fact, why have governments, despite their alleged concern over climate change, concentrated on funding research in the physical sciences to investigate the subject? Are these sciences driving policy — or vice versa?

Taking global warming seriously requires giving attention to issues such as the choice of fuels and technologies, energy pricing and investments¹. High economic stakes are involved. As a result, both the climate treaty and its underlying scientific debate have become swept up in global energy politics. The responsibility given to science is great — perhaps too much so for institutions increasingly

dependent on 'soft' contract research income.

Scientists naturally prefer to experiment with mathematical models of the Earth's various systems free of responsibility for policy². Uncertainty is their security. Indeed, some can already be seen withdrawing from policy involvement. For

underpin future national policies.

Much of the answer lies in shifts in the energy market in the 1980s. During this decade, both the Chernobyl accident and cheap fossil energy challenged forecasts of energy demand, and invited the involvement of energy interests in global politics. Energy prices generally collapsed in the middle of the decade and have remained low. This reversed the situation of the 1970s, when both the expansion of nuclear power and major research and development efforts on renewable energy technologies created major institutional interests. By the 1980s, these institutions found themselves under threat, and therefore began lobbying via well-established channels inside governments, leaving green rhetoric to the media, environmental groups and UN agencies.

Global warming can therefore be said to have gained its political relevance less from scientific evidence and speculation than from the unexpected collapse of fuel prices, which recreated an earlier world of cheap energy. A 60 per cent drop in oil prices occurred only months after scientists had made a sweeping statement on the possible dangers arising from growing fossil fuel consumption. The 'green' energy losers consequently tried to 'capture' the expected regulatory process, while coal (and to a lesser extent oil, for which substitution was harder) became the main villains.

So far, the oil industry, rather than nuclear power, has been the major winner. With expensive European coal likely to disappear altogether, and the former Soviet Union opening up its resources to the West, natural gas is replacing both coal and nuclear power in unregulated markets. Gas has become the 'rational' choice for generating electricity, with lower fuel and labour costs, reduced emissions, and the added bonus of avoiding further investments in costly acid rain abatement technologies, such as flue gas desulphurization.

Where this strategy aroused protests, the greenhouse effect was cited in justification. When the collapse of oil prices reduced tax revenue, the privatization of electricity became even more desirable. Individuals and firms were urged to invest in energy-saving measures, if only to reduce cost increases. 'No regret policies', rather than precaution, became the response to scientific uncertainty. As a result, policies could be rationalized as

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Demonstration against the UK tax on heating bills— changing social attitudes make more difference than defining energy problems.

others, including the chairman of the IPCC, global warming has become the justification for a crusade against materialism and for a 'new organizing principle' — the preservation of the Earth. Yet global warming could not have entered international politics without the support of influential voices from the scientific community.

How and why did scientists create public concern in the first place? And why was this taken up — far too rapidly for many scientists — not only by environmentalists, but also by governments? The political energy needed to turn a research topic into a treaty with major implications was generated surprisingly quickly, even though it can still be argued that the treaty does little more than codify the research, data collection and monitoring needed to

environmental, even though adopted for other reasons. The nuclear lobby had expected to reap a greater reward from its early investment in research; it now joined the chorus calling for energy prices to 'internalize' the external costs of fossil fuels, inviting subsidies for itself, or taxation for competitors — preferably both.

The British government, for one, discovered that, by changing the fuel mix, air-pollution control was feasible with the support of industry, leading to lower generating costs. Environmentalists could be accommodated, as well as vocal lobbies supporting energy efficiency and renewables. Still other groups again wanted to sell cleaner technologies to industrializing countries — and then promote privatization in order to make profitable investments — or demanded constraints on private mobility. Supporters of the idea of ecological modernization saw climate-change policies as steps towards sustainable development. Vast amounts of data on 'global change' would not only serve research, but also make the Earth's surface more visible for all who could afford to use the new Earth observation maps.

Global warming was also attractive to politicians because of its value to both critics and supporters of deregulation policies. In Britain, for example, separate groups pursuing commercial interests, foreign policy goals and domestic politics each discovered their own uses for the warming hypothesis. For some, the opportunities included the pursuit of global scientific research agendas, for others, the enhancement of bureaucratic power at home³. Some saw the possibility of thwarting the ambitions of the European Commission (EC) in Brussels by making sure that decision-making powers remain outside EC forums in the hands of less efficient — and poorly financed — UN bodies.

Calls for environmental regulation were generally attractive to environmental bureaucracies. Having been created in the early 1970s, these had remained relatively weak, and often closely tied to research in the natural sciences. Global warming allowed national bodies to expand their influence — both domestically and internationally — while organizations such as the UN Environment Programme (UNEP) and the World Meteorological Organization (WMO) have been able to attract considerably more attention and funding. Furthermore, beleaguered national politicians gained a world stage on which to indulge in global green rhetoric without, as yet, having to face issues of domestic implementation. A climate treaty became desirable to many groups — but not major measures to reduce emissions.

We can now ask how, and why, scientific institutions initiated the debate, and how they subsequently responded to its

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Wind energy — beneficiary of the energy crisis and of the global warming debate. These are Mitsubishi turbines under construction at Llandinam in mid-Wales.

politicization. In 1985, a group of research scientists, ecologists, climate and energy demand modellers mainly from Sweden, Canada, the United States and the United Kingdom met in Villach, Austria. UNEP, WMO and the International Council of Scientific Union (ICSU) sponsored the conference, which also attracted support from major environment/energy research bodies such as the International Institute for Applied Systems Analysis, Harvard University and the Stockholm Environment Institute.

Only two government scientists attended this 1985 symposium, one from the UK Meteorological Office, the other from the US Department of Energy⁴. The gauntlet thrown to the world was therefore largely European, the main challengers being contract research institutions with interests in the carbon cycle and climatic variability.

The Villach conference drew the attention of policymakers to the dangers of global warming, and the need for more climate related research. ICSU had just finalized its International Geosphere-Biosphere Programme (IGBP). Developed in the United States during the early 1980s, and adopted by the ICSU in 1986, it would address the problem jointly with the World Climate Research Programme (WCRP), which had been set up in 1979, and was also seeking new tasks⁵.

Warming was a high but uncertain risk. It was also expected to happen soon if — as energy demand forecasters predicted — CO₂ concentrations were to double within decades. This assumption has been incorporated into all existing models and adopted by IPCC, with the United States and the Netherlands providing emission scenarios for the first report in which the dates for such doubling were fixed.

In this context, stringent regulations would be needed to prevent warming, opening up energy markets to 'green' technologies (including fast breeders). Past CO₂ increases made warming intuitively more credible, although cynics pointed out that model predictions can always be manipulated. Several policy measures were recommended, including a

global legal instrument requiring countries to reduce the emission of greenhouse gases, particularly CO₂.

The Villach conference had been organized by individuals who formed the core of the Advisory Group on Greenhouse Gases (AGGG), a small panel agreed at the conference and meant to advise sponsors, as well as governments. Consisting only of independent (non-government) scientists, it remained a major influence on UN-related activity until 1990.

The message of the Villach meeting spread rapidly, backed up by the Brundtland Report and its message of sustainable development. Diplomatic circles had already been alerted by individuals such as the former British diplomat Sir Crispin Tickell, who studied climate change at Harvard while on secondment in the late 1970s, and has since maintained close links with scientists in Europe and the United States. In the mid-1980s, Tickell helped to persuade the British Prime Minister Margaret Thatcher that global warming was an issue that deserved both her attention and generous funding. He appears to have been equally persuasive with the EC, the United Nations, the British Foreign Office, and ICSU.

With the support of environmentalist networks pursuing their own goals, the new global threat soon flourished. The public became alarmed. Many governments, especially those (such as the United States) who depend on fossil fuel, did not like AGGG, but preferred the WMO, which they are able to influence directly through national representatives.

The WMO decided to 'capture' the AGGG. In 1987, its executive proposed to set up a small body of experts from countries strong in climate-change research, including a small number of people representing other nations and international scientific bodies such as ICSU. The IPCC evolved from this idea during the subsequent politicization of the issue. Formally set up in autumn 1988 by the WMO and UNEP, its activities are controlled by government appointees who, in practice, tend to be influential government scientists with connections to the WMO and other international bodies.

Effective IPCC leadership fell to a small group of committee members, most of whom had strong links to NASA, as well as large national laboratories and meteorological offices. The research networks supporting two of the three working groups which were set up in 1988 already existed as parts of World Climate Research Programme. The first of these, for example, the science assessment group, is managed through the UK Meteorological Office, and funded by the Department of the Environment. The second working group, which is looking at the impact of climate change, was in the hands of the Russians, strongly supported by Australia, but has recently found a home within NASA, with a chairman who advises the US president.

Initially, government officials and representatives of various pressure groups formed the third, less permanent group. This was asked to produce realistic 'response strategies' on the basis of the 'scientific' inputs from the other two groups — even though these were not yet available. Impacts and responses were combined in 1993, while a new group was set up to look at 'cross-cutting' aspects, with contributions from a wider range of social sciences.

In 1990, the WMO informed policymakers that they should consider the consensus views of the scientific community as 'timely assessments' of the climate threat⁶. The IPCC was made available to them as a 'single established source' for information on this subject, and while the committee's task would not be to define or set up research programmes, its recommendations would be of value to the agencies responsible for research — ICSU, WMO and UNEP.

The IPCC 1990 science assessment was produced by Britain, with major assistance from a small group researching ozone science inside NASA. It provided political legitimacy to a national research agenda coordinated globally by ICSU, WMO and UNEP, and already defined — in the context of the warming threat by the ICSU's Committee on Problems of the Environment, SCOPE — for the 1985 Villach meeting⁷.

IPCC reports summarize the findings of its research networks, with drafting and editing carried out by small groups of lead authors. These tend to be selected by the science managers, who also draft IPCC 'policymakers' summaries — themselves important political documents and skilful exercises in scientific ambiguity⁸. A widely praised 'scientific consensus' of 1990, for example, which assessed the prospects of climate change, used language which simultaneously allowed Greenpeace to call for a target of reducing emissions by 60 per cent, and the UK Treasury to conclude that no action was needed until more scientific certainty was available —

each citing the same source.

A second assessment exercise is in preparation. But the credible regional predictions and emission models so urgently desired by policymakers are not yet available. Global experimentation continues.

So far, there has been little redirection of global energy policy to take account of the potential threat of global warming. This failure could be blamed on the scientific community, on falling energy prices or on politicians. Only if it could be shown that scientists intentionally created a more powerful threat than the evidence suggested, could this lack of response be considered wise. There is some evidence for this⁹, although a responsiveness of 'science' to both internal need and external pressure is surely not surprising.

But, given both the energy politics of the mid-1980s, and the large number of non-scientific actors involved, the rapid politicization of the climate debate occurred in a context of scientific ambivalence, influenced by forces beyond the control of science. Global policy on global warming is emerging from untidy political processes — not through technocratic design. Given the scientific uncertainties that still exist, this may well be for the best.

IPCC has undoubtedly boosted research activities in countries with strong atmospheric, oceanographic and space research capacities. It has also stimulated mathematical modelling of some global macro-economic factors, as well as of carbon emissions related to fuels, demand and technologies. Most of the OECD countries, as well as those in the EU, now possess major global environmental change research programmes which also consider institutions and broader social questions.

There have been rewards to those institutions in advanced countries who, with the help of environmentalists, have been able to link science policy to at least symbolic commitments to the global environment. Furthermore, uncertainties in scientific advice have served the interests of research by reinforcing calls for extra research funds.

But political battles over the knowledge base gave also grown. This may explain the withdrawal of science into adopting a more neutral position on appropriate policy responses than appears to be the case between 1985 and 1992, incidentally the preparatory phase for the Rio conference and for beginning the implementation of the IGBP. Now that the authority of science has been weakened, claims to both policy relevance and truth — as well as the capacity of the natural sciences to deliver policy advice that can be quickly implemented — can legitimately be questioned. The creation of knowledge at increasing rates by a few individuals

engaged in a political game, one of whose primary goals is to obtain funding for their own work, must influence the way that environmental problems are defined and prioritized.

Resistance to universalist and politically 'neutral' claims made by Western politicians in the name of science appears to be growing, while appeals to ethics and issues such as fairness, equity and 'social exclusion' are growing stronger.

In 1971 the American climatologist R. A. Bryson voiced 'a sneaking suspicion' that calls for more data and monitoring prior to the Stockholm Conference were "mostly for the care and feeding of big computers"¹⁰. He speculated that it would be more appropriate to monitor parameters with a more direct impact on human welfare. In Rio in 1992 it was agreed that the world should be concerned that over 12 million children a year die from preventable diseases. But the types of science and technology that promise commercial competitiveness and political power do not address such issues.

Under pressure, even scientists will deliver what their paymasters prefer to hear. Honest science may be less useful, in the short term, than relevant science, a lesson that has to be continuously relearned. Policy-related advisory networks need to become more sophisticated and less self-serving — and policymakers to develop broader decision-making structures — to prevent this from happening. This cannot mean listening to green or energy lobbies to the exclusion of all others. Advisers and decision-makers need each other, but knowledge funded by soft money and created under conditions of dictated relevance and competitive bidding is surely unlikely to inspire the degree of trust upon which wise policy must be based. □

Sonja A. Boehmer-Christiansen is a senior research fellow at the Science Policy Research Unit at the University of Sussex, Brighton BN1 9RF, UK.

1. *World Energy Council Journal* (July, 1993).
2. *AMBIO* 23(1), (February, 1993).
3. Boehmer-Christiansen, S.A. *Env. Politics* (in the press).
4. Report of the Second International Conference on the Assessment of Carbon Dioxide and Other Greenhouse Gases in Climate Variations and Associated Impacts, Villach 9-15 October 1985 (WMO 661, Geneva, 1986).
5. *The World Climate Programme 1992-2001* (WMO 762, Geneva, 1992).
6. Jager, J. & Ferguson, H. L. (eds) *Climate Change: Science, Impacts, Policy: Proceedings of the Second World Climate Conference* (Cambridge University Press for WMO, 1991).
7. Bolin, B. et al. (eds) *The Greenhouse Effect: Climatic Change and Ecosystems* (SCOPE 29, Wiley, Chichester, 1986).
8. Houghton, J. et al. (eds) *Climate Change: The IPCC Scientific Assessment* (Cambridge University Press, 1990).
9. Boehmer-Christiansen, S.A. *Energy and Environment* 4, (1993).
10. Bryson R.A. *The Environmental Future* (ed. Polunin, N.) 165 (Macmillan, London, 1972).

ACKNOWLEDGEMENTS. The author was partially funded by the ESRC.

The poor quality of random numbers

After decades in which the army of computer simulators have used machine-generated random numbers with confidence, even aplomb, some generating algorithms turn out to be frayed at the edges.

WHAT exactly is a random number? Many senses in which that question may be understood evidently make no sense. For example, it is nonsensical to ask in the abstract whether "3" is a random number. For one thing, it cannot be described as random without specifying the range of numbers from which it has been chosen. For another, even if it has been chosen by some random process (from among, say, the integers between "1" and "10" inclusive), that will be entirely irrelevant unless the uses to which the number is put are specified.

In this spirit, one might set out to calculate the scattering cross-section of a moving billiard ball by another at rest, building a simple computer program to simulate the typical collision and then assigning trajectories chosen at random to consecutive impactors. Then one becomes an important user of random numbers. The same result can be obtained more easily with a little calculus, but not all calculations are that amenable to analytical treatment. In any case, what matters in such a context is that the numbers should be random with respect to each other. By now, there are many procedures, known by the generic name of Monte Carlo, that follow these lines and which are major users of the commodity called the random number, which has to be manufactured somehow.

That is not child's play. Here is an account (by Robert M. Ziff, *Phys. Rev. Lett.* **69**, 2671; 1992) of one technique:

"...a total of 6×10^{12} random numbers were generated, requiring a few months computing time on about a dozen computer workstations (Apollo 425, IBM R/S 6000) running simultaneously."

Random numbers made in ways like this are not really random, but are the products of computational schemes or algorithms. So, in the last resort, a set of random numbers cannot be entirely free from non-randomness. Everybody in the trade knows that, and tries to arrange that calculations are so designed that the non-randomness will not embarrassingly become apparent. But that seems not always to be possible.

Indeed, for the past two years there has been a zephyr (hardly a gale) of excitement that a calculation by Alan M. Ferrenberg, D. P. Landau and Y. Joanna Wong, from the University of Georgia at Athens, appears to have produced erroneous results in circumstances that cast doubt on the quality of the random numbers they used (*Phys. Rev. Lett.* **69**, 3382–3384; 1992). This was at the time more than a mere cloud on a blue horizon; NATURE · VOL 372 · 1 DECEMBER 1994

some of the random-number generators used go back for twenty years, and had not previously been found wanting.

The test system used by Ferrenberg and his colleagues was the familiar Ising lattice, a two-dimensional square lattice supposed occupied by classical spins pointing in one direction or the opposite. The Ising lattice is a model for several kinds of critical phenomena. The usual procedure is to simulate an infinitely large lattice by supposing it to be broken into finite rectangles and imposing the same boundary conditions at the edges of each rectangle making up the whole plane.

The object of the exercise is to calculate the thermodynamic properties of such a system, which is a matter of calculating the energy of particular configurations of spins and, from them, the partition function (a function of the temperature). The snag is that, except when the lattice is small, the enumeration problem is huge. So people have fallen back on Monte Carlo methods to arrive at the equilibrium state of the system in one go. The simple way is to invert a single spin, chosen at random, and to calculate the effect on the thermodynamic properties of flipping the spins of its nearest neighbours, their nearest neighbours, and so on. That state in which a single flipped spin makes no overall difference is the equilibrium state.

The practical difficulty is that the simulation methods become extremely slow near the critical temperature at which the lattice is transformed from the ordered to the disordered state. So Ferrenberg and his colleagues were anxious to try a novel method due to Ulli Wolff from the University of Kiel, in which the approach to equilibrium in an Ising lattice involves switching whole clusters of spins simultaneously. This was reckoned to be much faster, but Ferrenberg found that it gave "systematically incorrect results". They knew that because the simple two-dimensional Ising lattice is exactly calculable.

Three Finnish workers have now shown that the errors found by Ferrenberg and colleagues are indeed traceable to the random numbers used (Vattulainen, I., Ala-Nissila, T. & Kankaala, K. *Phys. Rev. Lett.* **73**, 2513–2516; 1994). They test their conclusions with a variety of generators of random numbers. An account of the manufacture of random numbers follows.

Everything is done by computer, so that numbers are most conveniently represented as binary digits, preferably as many as there are bits in one computer byte. Evidently an iterative process, in which one random number is generated from others already

defined, is simplest when there is a need to make a lot of them. Suppose that the numbers each consist of 32 binary bits ("0"s and "1"s) and that there are already n of them in being, say $x_1, x_2, \dots, x_n$. Then one could imagine forming x_{n+1} by using some simple computer operation to combine some specified pair of the pre-existing numbers together. One of the simplest computer operations is the "OR" or "XOR" operator which, operating one bit at a time, turns one 32-bit number into another. That is the basis of what is called the generalized feedback shift register (GFSR) generator of random numbers, in which $x_n = x_{n-p} \otimes x_{n-q}$, where $\otimes$ means the OR operator and p and q are integers.

There are more elaborate ways of making random numbers, but to the extent that they all depend on algorithms, they are all subject to non-randomness. But it is clear that the GFSR method requires at least p or q numbers (whichever is the larger), whence the benefits of the algorithm $x_n = (16807 \times x_{n-1}) \bmod (2^{31}-1)$, apparently variously known as GGL, CONG and RAN3.

The Finnish group has repeated the simulations that caused a stir at the end of 1992, discovering that discrepancies do indeed depend on the random-number generator. They have gone on to devise two neat physical tests of the randomness of random numbers. One, called the "random walker test", would have the random numbers make a point move between one of the four quadrants in a square, in the expectation that it would occupy each of the four sub-squares equally often. In reality, they find that some favourite algorithms, especially the GFSR generators with the larger of p and q equal to 31, 250 and 521, as well as RAN3, all fail the simple test.

The other test is even simpler. Suppose the random numbers are scaled to fit between 0 and 1, and that they are taken together in blocks of n as they come off the battery of perpetually running workstations. Take the average of each block and assign the number 0 if the average is less than 0.5 and, otherwise, 1. There should be an equal number of the two digits, and the probability of departures from equality can be obtained by simple statistics. Again the same random-number generators fail.

Primarily, all this will be a warning to the Monte Carlo community, but others will be surprised that the errors have come to light so late in the day, and that so little should be known of their origin. Perhaps that is something with which to occupy the long Finnish winter nights ahead.

John Maddox

Events after the events

Paul Weissman

WHAT really happened when the fragments of comet Shoemaker-Levy 9 slammed into Jupiter this July? At the first two post-impact scientific gatherings*, observers and theorists were overwhelmed trying to explain the complex sequence of events that were seen each time a cometary fragment dived into the atmosphere at 60 km s^{-1} . Some answers are beginning to emerge, but they are dwarfed by the many more questions being raised.

Some of the most important questions are the same as before the impacts occurred: how big were the cometary fragments and how deep did they penetrate into the jovian atmosphere? Other questions are entirely new: how did Earth-based observers see 'around the corner' and actually observe the impact events on the nightside of Jupiter, as seems to have occurred? What was the composition of the huge dark plumes of debris thrown out of the impact sites, forming clouds larger than the Earth in the jovian stratosphere?

The key to answering many of these questions may be new data just returned from the Galileo spacecraft *en route* to Jupiter, which had the only direct view of the cometary impacts on the nightside of the giant planet. Galileo was still about 240 million km from Jupiter when its instruments recorded the impact events in July. Because of Galileo's damaged high-gain antenna, data have trickled back to Earth at a mere 10 bits per second since then.

A joint observation of the impact of Shoemaker-Levy 9's G fragment, one of the largest, by Galileo's near-infrared mapping spectrometer, photopolarimeter-radiometer, and ultraviolet spectrometer, was described by Robert Carlson and Terry Martin (Jet Propulsion Laboratory) and Wayne Pryor (Univ. Colorado). They observed a 7,500 K fireball perhaps 10 km in diameter, appearing at the top of the jovian clouds. The infrared data at wavelengths from 1.8 to $4.4 \mu\text{m}$ show the fireball growing at about 2.2 km s^{-1} over more than a minute, cooling adiabatically in the process. The photopolarimeter data for the same event show the fireball fading out after about 35 s, shorter because the instrument was observing at a shorter wavelength, $0.945 \mu\text{m}$, where the cooling fireball disappeared more quickly.

The profile of the G event recorded by the photopolarimeter is the same as that instrument observed for H and L, both large impacts. A very similar profile was

also reported for the K event, another of the largest impacts, by Clark Chapman (Planetary Science Institute) of the Galileo CCD camera team. Observing at $0.882 \mu\text{m}$, the camera saw a bright flash lasting about 5 s, which then faded and brightened again in about 10 s, and remained visible for another 30 s. But for one of the smaller impact events, N, the camera saw a bright flash lasting about 5 s,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 1 Ejecta plume from the impact of fragment G on Jupiter, 18 July 1994. Universal times and wavelengths at which the images were taken are (top to bottom) 7:33, methane; 7:38, red; 7:41, green; 7:44, blue; and 7:51, violet. The impact site is on the nightside of the planet but the plume of ejecta is visible because it rose into the sunlight, to an altitude of 3,300 km. The top image is very close in time to the detection of the impact by instruments on the Galileo spacecraft, and appears still to be in the shadow of Jupiter; it may show light from the explosion fireball reflected off late-infalling cometary dust. In the later images the ejecta plume expands, cools and then falls back onto Jupiter's stratosphere.

followed by a much dimmer fireball lasting only about 10 s. A similar five-second initial flash was seen for the W event, but the subsequent data have not yet been returned to Earth.

The Galileo investigators' interpretation of these observations is that the first 5-s flash is the comet fragment entering the atmosphere and beginning to burn up like a meteor or bolide. As the fragment plunges into the atmosphere, shedding energy and material along its path, it superheats a column of jovian atmosphere. This column explodes instantly and

creates the intense fireball seen by the Galileo instruments for the next 10–60 s. Curiously, all seven bolide entry flashes seen so far are of about the same brightness, whereas the magnitude and duration of the fireballs appear to reflect the relative sizes of the impactors. More Galileo data on the G, R and W impacts will be played back during December 1994 and January 1995.

Although there are still discrepancies in the timing of events, it now appears that the combined bolide/fireball events seen by Galileo correspond to the first of two precursor flashes that Earth-based infrared observers often saw before each impact site rotated into their view. But how did ground-based observers see these events when they were behind the planet? One speculation is that the flash was reflected off a dust trail following each comet fragment, or a cloud of high-altitude debris deposited by dust which preceded each fragment into Jupiter's atmosphere. Alternatively, observers may have been seeing the beginning of the fragment's glowing, meteor-like interaction with the thin upper atmosphere.

The second precursor flash seen by the Earth-based observers is believed to be the plume of high-speed ejecta from the impact rising into the sunlight above Jupiter's limb. Images from the Hubble Space Telescope show the development of the plume over a period of about 20 minutes each for four of the impacts (see Fig. 1). Emerging at velocities of 10 to 15 km s^{-1} , the ballistic ejecta forming the plumes fell back on the upper stratosphere of Jupiter, heating a huge area of atmosphere, typically about the diameter of the Earth or larger. As they rotated into view from the Earth, these hot clouds of debris produced the immense infrared brightenings seen by ground-based observers.

But what made up the huge, dark clouds of debris, easily visible in even amateur telescopes from Earth? Presumably, each comet fragment was vaporized on impact and its constituent molecules dissociated, together with a considerable amount of jovian atmosphere along the explosion path. Large amounts of sulphur compounds, carbon monoxide and water were observed in the plumes. Gordon Bjoraker (NASA Goddard Space Flight Center) and colleagues observing with NASA's Kuiper Airborne Observatory reported a quantity of water in one impact site equivalent to a 1-km-diameter sphere of ice. The amount of sulphur was so great that it must have had a jovian source, presumably the clouds of ammonium hydrosulphide in Jupiter's atmosphere. But the sources of the carbon, oxygen and hydrogen were much less clear. Bjoraker believes that the high abundance ratio of water to methane and the high temperature of the water in the plumes means that the source was the cometary fragments

*Hypervelocity Impact Symposium, Santa Fe, New Mexico, 17–19 October 1994; AAS Division for Planetary Sciences meeting, Bethesda, Maryland, 31 October–4 November 1994.

themselves. In addition, emission lines from various metals were observed and these were almost certainly from the comet fragments.

The confusing chemistry even led some observers to suggest that Shoemaker–Levy 9 was actually an asteroid rather than a comet. Set against that, however, is the persistent coma that the comet fragments displayed from their discovery right up until hours before impact. In addition, orbital statistics make a cometary origin for Shoemaker–Levy 9 far more likely. Detailed modelling of the chemistry of the fireballs will be needed to unravel the spectroscopic evidence. An aid to solving the problem will be the measurements of jovian atmospheric composition and structure, which will be made by the Galileo atmospheric entry probe when it plunges into Jupiter on 7 December 1995.

An interesting feature of the impact plumes reported by Heidi Hammel (Massachusetts Institute of Technology) is that they all appear to be much the same height, about 3,300 km, regardless of the size of the impactor that generated them. This confounds the impact modellers who have tried to use the heights of the plumes to estimate the diameters of the comet fragments. It seems that plume height is not a very meaningful discriminator.

There are some points that the modellers do agree on. The simple concept of each cometary fragment entering the jovian atmosphere as a meteor and then exploding in a sudden, suicidal point explosion is not correct. As noted above, when the comet enters the atmosphere it sheds material and energy along a column of superheated gases which it drills in the jovian atmosphere. This entire column explodes, somewhat like a line charge of explosives, but with varying effects because of the irregular pattern of the energy deposition and the increasing atmospheric density as the comet fragment plunges through many scale heights of atmosphere. Most theorists also agree that the comet fragments seem to have deposited a large amount of energy high in the atmosphere, above Jupiter's cloud decks.

At that point, the consensus stops. Modellers Kevin Zahnle (NASA Ames Research Center) and Mordecai-Mark Mac Low (Univ. Chicago) believe that the comet deposits energy high up because it is small, only 0.5–1.0 km in diameter, with a density of 0.5–1.0 g cm⁻³. Their model predicts that the comet penetrates the ammonia and ammonium hydrosulphide clouds on Jupiter, but does not reach the

water clouds at about the 5 bar pressure level in the atmosphere. The cometary debris is then ejected back up the column in a high-speed jet that forms the ballistically emplaced debris clouds. This model explains the large amounts of sulphur but lack of jovian water observed in the ejecta plumes (although as noted above, the source of the water in the plumes is not clear). Also, the model agrees with predictions of the size of the impactor fragments based on modelling of the tidal

the Hubble Space Telescope.

Yet another model, by Toshiko Takata and Thomas Ahrens (California Institute of Technology), finds an intermediate impactor size, 2.1 km in diameter, for the largest events. Like the others, it predicts huge plumes emerging from the impact sites. Using three-dimensional simulation, Takata and Ahrens also find the comet fragment penetrating deep into the atmosphere, below the water clouds. But in their model both the jovian and cometary water fail to rise high enough in the atmosphere to be observed, and remain obscured by the plumes.

NASA/Hubble Space Telescope

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 2 The face of Jupiter before it was hit by the comet fragments (top, taken on 15 July 1994) and after (bottom, taken on 23 July 1994, immediately after the impacts ended). Each of these two colour composite maps was assembled from images taken on at least five Hubble Space Telescope orbits. The lower map shows that although the impact sites are fresh, Jupiter's winds are already smearing them, principally east–west, along latitude lines.

disruption of the original comet (J. V. Scotti and H. J. Melosh, *Nature* **365**, 733–735; 1993, and E. Asphaug and W. Benz, *Nature* **370**, 120–124; 1994).

At the other end of the modelling spectrum are simulations by Mark Boslough and David Crawford (Sandia National Laboratories), who find that the impactor had to be large, up to 3 km in diameter, so that its large cross-section would result in substantial energy deposition at altitude. Boslough and Crawford find that the larger comet fragment plunges deeper into the jovian atmosphere, well below the water clouds, but that the atmospheric pressure at that altitude contains the explosion and prevents it from expanding back up the superheated column. In effect, Jupiter's atmosphere swallows up most of the energy of the explosion, and most of the cometary material and jovian cloud material remains at depth where it is invisible to Earth-based observers. This model uses a three-dimensional simulation rather than Zahnle and Mac Low's two-dimensional simulation, so the physics should be more complete. However, they fail to explain why large amounts of sulphur were seen in the impact plumes by

It may be possible to reconcile the different modelling results with a relatively small impactor if each comet fragment was a 'rubble pile' of smaller icy-conglomerate sub-fragments. Tidal forces would cause the rubble pile to begin to disperse as it passed into Jupiter's Roche limit, although there would not be enough time for the fragments to spread very far. But it might just be enough to give a large effective cross-section, resulting in considerable energy deposition at high altitudes. For now, it is not clear if this actually occurred.

The morphology of the debris clouds (see Fig. 2) may also be a clue to how deep the impactors penetrated, and therefore how large they were.

Each impactor created a central dark cloud of debris where it struck Jupiter, but only the largest impactors created broad, dark clouds of ballistic ejecta out to distances of 6,000 to 8,000 km from the impact sites. Because all the impactors appeared to generate equally high plumes, why were the ejecta from only the largest impacts visible? Mac Low speculated that only the largest impactors dredged up sulphur compounds from the jovian ammonium hydrosulphide cloud deck, and it was the various sulphur compounds that made the debris clouds easily visible.

Meetings on the Shoemaker–Levy 9 impacts are planned for next March at the European Southern Observatory in Garching, and for May at the Space Telescope Science Institute in Baltimore. By that time, comparison and coordination of the many observational data sets, advanced modelling efforts and the remaining Galileo data may come together to provide answers to the many questions that have been raised so far. □

Paul Weissman is in the Earth and Space Sciences Division, Jet Propulsion Laboratory, Pasadena, California 91109, USA.

In search of a satiety factor

Timothy J. Rink

PHYSIOLOGISTS have long postulated that, when a mouse or man overeats, the resulting extra fat somehow signals to the brain that the body is obese and makes it eat less and burn more fuel. On page 425 of this issue¹, Friedman and colleagues report the cloning of the murine *obese* (*ob*) gene, the normal product of which may be a hormone that is secreted by fat cells and controls body weight.

Forty four years ago, a genetic defect (the first of several) was identified in mice which leads to them becoming morbidly obese when homozygous for the mutation (see cover). The gene concerned was *ob*. For the past eight years, starting with the interbreeding of hundreds of mice, there has been a concentrated research programme to identify *ob* and it has now culminated in the cloning of the gene. Friedman's team has isolated a complementary DNA that is expressed specifically in fat cells. The predicted sequence of the gene product has 167 amino acids and the characteristics of a secreted protein. It is 84 per cent the same in mice and humans, and a homologous sequence was found in several other mammals, and chicken and eel (but not fruitflies). It seems that mice homozygous for the *ob* mutation fail to make and secrete the normal protein product — making *ob* protein a putative 'satiety factor'.

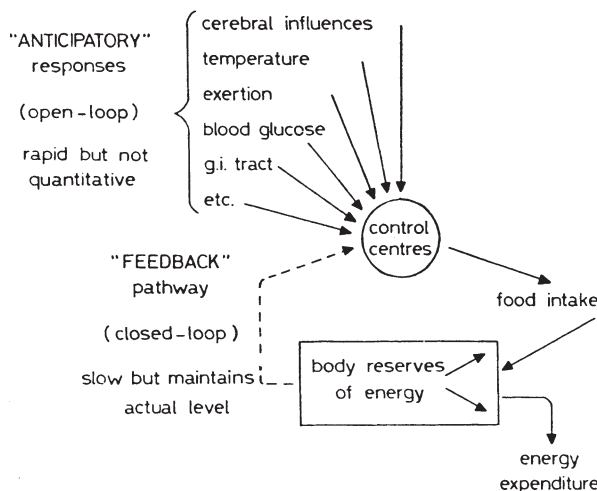
These studies form a pivotal step in developments stemming from a fascinating *ménage à trois* of physiology, classical genetics and molecular biology. A number of physiologists (including myself) have inferred that, in the face of fluctuating food intake and energy expenditure, the precision of body-fat regulation (within ± 1 per cent over many years) requires a powerful, slow feedback pathway; and that total fat mass is the variable that is sensed and regulated. This 'lipostat model' of Kennedy² has been discussed by Hervey³, Mrosovsky⁴ and Weigle⁵, among others (see figure).

The following points encapsulate key data from many decades of work on body-weight control. Most of the experiments were done in rodents, but there are enough parallels with other species, including humans, to infer that the principles apply to mammals generally.

■ The hypothalamus seems to be the main control centre for satiety and energy expenditure. Localized damage to this region of the brain causes an obese phenotype similar to that of *ob/ob* mice; on the

other hand, appropriate stimulation of the hypothalamus reduces eating and increases energy expenditure.

■ Rats forced to overeat lay down excess fat, but when offered a normal diet they eat less until normal body weight is restored. The half times (several days) for weight reduction and normalization of food intake are similar.



Hypothetical model of the mechanism that regulates energy balance in mammals, as depicted in 1969. (Reproduced from ref. 3.)

■ Removal of a substantial mass of fat is followed by extra eating and an increase in the remaining fat stores.

■ Overfeeding of one of a pair of parabiotic mice (mice which are surgically joined and have some interchange of blood) reduces food intake and induces weight loss in the partner⁶, apparently because of the transfer of a circulating hormone.

■ A ventro-medial lesion to the hypothalamus of one of a parabiotic pair leads to obesity in that mouse, but results in reduced food intake and loss of adipose tissue in the other animal. Again, some satiety factor from the obese animal appears to be transferred.

■ If an *ob/ob* mouse is surgically joined to a normal animal, it eats less and gains less weight⁷. This finding fits in with the idea that the *ob/ob* phenotype results from lack of the proposed satiety hormone, which is supplied by the normal animal in the parabiotic pair.

■ Mice homozygous for another mutation, *db*, are also obese. The *db/db* phenotype appears to reflect a defect in the action of *ob* protein, perhaps a receptor defect. This conclusion is in line with data from parabiosis experiments in which a normal mouse, paired with a *db/db* mouse, rejects food and dies of starvation⁷. Consistent with the results of these experiments, *ob/ob* mice made para-

biotic with *db/db* mice reduce their food intake and lose adipose tissue. So the *ob/ob* mice appear to respond to the putative excess of *ob* protein produced by their *db/db* partners.

In the face of this and Friedman's new evidence, it is difficult to reject Weigle's conclusion that "the adipose tissue satiety signal must be one or more currently unidentified circulating molecules with a $t_{1/2}$ in excess of 40 minutes (ref. 5). Hervey³, however, had suggested a slightly more complex 'dilution hypothesis', whereby a lipophilic factor, possibly a steroid, is produced and disposed of at a constant rate, and is diluted into the fat store; the fatter the animal or person, the lower would be the factor's plasma concentration. If this factor increased food intake and decreased energy output, then it could appropriately alter them in response to increasing or decreasing fat mass. But in this form, the dilution hypothesis is actually flawed; the disposal rate needs to be a function of fat mass for the system to provide the requisite slow feedback pathway. In this model the *ob* mutation could cause inadequate disposal, and hence accumulation of an 'eating' factor, while the *db/db* mutation could cause excess production of the factor. The parabiosis experiments could all fit into this model

provided that the 'eating' factor had a long biological half-life.

Friedman's data make Weigle's concept of a 'fat-derived satiety factor' the more promising hypothesis. Indeed, one report points to the existence of a satiety factor, extractable from adipose tissue of overfed rats, that suppresses food intake for over 12 hours⁸. The *ob* gene apparently encodes a protein secreted by fat, and mutations apparently prevent translation or expression of the gene. The *ob* protein is thus likely to be a fat-derived molecule with a long half-life, which acts on the hypothalamus to exert long-term, overriding control of appetite and, most likely, of fuel storage and energy expenditure.

Some cases of morbid obesity in humans may reflect a homozygous condition analogous to the *ob/ob* mouse. With the availability of Friedman's data, investigations of such cases should be relatively straightforward. The more common forms of obesity may reflect sub-normal production of the *ob* protein. Body weight must stabilize at an excess level, one at which sufficient *ob* protein is produced to influence the hypothalamus to match food intake to energy intake. Dieting is likely to be only temporarily successful because weight loss will reduce the production of the satiety factor, and there will be an enduring drive to increase food intake

and reduce energy use until weight again stabilizes at the original obese level, with appropriate production of *ob* protein.

However attractive it may be to fit the new data on the *ob* gene into the lipostatic theory, we must remember that what Friedman and co-workers have actually come up with is the sequence of the deduced *ob* open reading frame. Much biochemistry and physiology has yet to be done to validate the concepts discussed above. We need to find out whether *ob* protein is actually secreted by adipose tissue; whether its circulating levels alter with total fat mass; and whether it acts as a long-term satiety hormone *in vivo* (if it does, excess *ob* protein should induce starvation in the face of abundant food — truly nature's slimming remedy).

Further key questions include the following. Do known hormones such as insulin regulate *ob* protein secretion? Where and what are the protein's receptors — are they in the brain, or also on peripheral tissues such as fat and liver? We also need to know how the long-term feedback pathway interacts with short-term regulators of appetite such as cholecystokinin and neuropeptide Y. Of course, we cannot expect one gene to be

the only controller, and it will be necessary to identify the genetic factors which lead some individuals to accumulate more fat than others even when they have closely matched diets and exercise regimens. Furthermore, how does *ob* fit in with other genes such as *agouti*? Agouti protein is expressed in adipose tissue and acts as an antagonist to melanocyte-stimulating hormone receptors⁹, but its involvement in obesity is not clear.

So much remains to be done. But it will not have escaped the notice of the authors or readers that cloning the *ob* gene may provide new and rational approaches to the therapy of obesity. □

Timothy J. Rink is at Amylin Pharmaceuticals Inc., 9373 Towne Centre Drive, San Diego, California 92121, USA.

1. Zhang, Y. *et al.* *Nature* **372**, 425–432 (1994).
2. Kennedy, G. C. *Proc. R. Soc. Lond. B* **140**, 578–592 (1953).
3. Hervey, G. R. *Nature* **227**, 629–631 (1969).
4. Mrosovsky, N. *Physiol. Behav.* **38**, 407–414 (1986).
5. Weigle, D. S. *FASEB J.* **8**, 302–310 (1994).
6. Harris, R. B. S. & Martin, R. J. *Am. J. Physiol.* **247**, R380–R386 (1984).
7. Coleman, D. L. *Diabetologia* **9**, 294–298 (1973).
8. Hulsey, M. G. & Martin, R. J. *Physiol. Behav.* **52**, 1141–1149 (1992).
9. Lu, D. *et al.* *Nature* **371**, 799–802 (1994).

PALAEOCLIMATOLOGY

New prospects in old bubbles

Kathryn M. Gregory

ANYONE who has panted and plodded up a mountain peak has felt the profound effects of decreasing atmospheric pressure with elevation. But what in the geological record might document this important variable? If palaeoatmospheric pressures could be estimated at given locations, these would be clues to the evolving elevation of landscapes, and also to the formation and composition of the early atmosphere. We can gain some idea from plants, which react indirectly. Temperature and enthalpy change with elevation, so vegetation type also changes. At first glance, one might suppose rocks to be impervious to atmospheric pressure. However, Sahagian and Maus, on page 449 of this issue¹, propose a clever method to estimate palaeopressure from bubble sizes in basalt flows.

Sahagian and Maus argue that the bubble distribution of a given basalt flow at emplacement is uniform, with a Poisson distribution of sizes. Several centimetres at the very top and bottom of the flow cool rapidly and preserve the original vesicle distribution. The modal size will be larger in the upper crystallized zone, because the size of a bubble of a given gas content varies as a function of the sum of overburden (flow thickness and density) and atmospheric pressure. Thus one only has

to measure the relative modal sizes in the very top and bottom zones to calibrate the size distribution to the effects of overburden. Atmospheric pressure then can be estimated.

If the method works, there are at least two exciting applications. First, if one knows the elevation of emplacement, for instance when the flow is at sea level, then one can estimate the palaeoatmospheric pressure. Such data could contribute to discussions about the composition of the Earth's early atmosphere and the carbon cycle; for example, whether cooler temperatures from a 'faint young Sun' were compensated by a greenhouse effect caused by high atmospheric carbon dioxide levels.

Second, if one assumes a certain atmospheric pressure at sea level for a given flow, then one can estimate the elevation. Palaeoelevation data are essential to provide realistic topography for atmospheric circulation models and to set limits on the vertical dynamics of continental evolution and deformation. Yet these data are notoriously difficult to assess. Currently the only reliable palaeoaltimeters that record absolute elevation are based on the foliar physiognomy method developed by J. A. Wolfe²; basalt vesicularity would be a welcome addition. Other methods used

to infer amounts of surface uplift include stream incision, uplift of geomorphic surfaces, increases in clast size or unit thickness, increases in erosion rates or exhumation, and displacement of geodetic benchmarks, but these are known to be problematic³.

In one foliar method, the mean annual temperature of an inland palaeoflora, derived from analysis of leaf morphology, is compared to that of a coeval coastal palaeoflora, and then a terrestrial lapse rate (the change of temperature with altitude) is applied to estimate elevation. This method has been criticized for relying on lapse rates, which may vary with global climate. However, another foliar method eschews lapse rates, relying on estimating elevation via palaeoenthalpy, and derives results similar to the previous method (C. E. Forest, P. Molnar and K. A. Emanuel, manuscript in preparation). Basalt vesicularity would provide a measure of palaeoelevation independent of biology, and would greatly expand the number of sites suitable for palaeoaltimetry studies. If basalt and floral studies are used in conjunction, workers could build high-quality uplift histories for a given orogen.

Such a data set could shed light on the Himalayan controversy. Some argue that uplift of the Himalayas caused Late Cenozoic global cooling⁴, whereas others argue that it was the other way round: Late Cenozoic cooling caused erosion of valleys and isostatic uplift of the high Himalayan peaks⁵. Resolving this debate is central to the understanding of how climate changes and how topography is built.

But there are caveats associated with a necessarily simplified view of magma dynamics and internal pressure variation. Sahagian and Maus sample seven 1.02–1.78-m-thick flows at a low and a high site on the volcanic peak Mauna Loa to test the resolution of their technique. In an ideal situation, where a single-event basaltic sheet flow is emplaced instantaneously at a given thickness, the basalt vesicularity method should work well. But work by Hon *et al.*⁶ suggests that such flows are less common than previously thought; they found that pahoehoe sheet flows from Kilauea characteristically propagate at thicknesses of 0.2–0.3 m. Within tens of minutes, the crystallized crust becomes thick enough to retain incoming lava on flat slopes, and the ensuing increase in hydrostatic pressure results in thickening, or inflation, of the

1. Sahagian, D. L. & Maus, J. E. *Nature* **372**, 449–451 (1994).
2. Wolfe, J. A. *Bull. U.S. geol. Surv.* **1964**, 1–35 (1992).
3. England, P. & Molnar, P. *Geology* **18**, 1173–1177 (1990).
4. Ruddiman, W. F. & Kutzbach, J. E. *J. geophys. Res.* **94**, 18409–18427 (1989).
5. Molnar, P. & England, P. *Nature* **346**, 29–34 (1990).
6. Hon, K., Kauahikaua, J., Denlinger, R. & MacKay, K. *Geol. Soc. Am. Bull.* **106**, 351–370 (1994).
7. Sahagian, D. L., Anderson, A. T. & Ward, B. *Bull. volcan.* **52**, 49–56 (1989).

entire sheet flow up to four metres.

Note that the pressure equation refers to thickness at the time the outermost zones crystallized (within 1–2 hours) not final thickness (attained tens or hundreds of hours later). Thus one would want to sample flows with as little post-emplacement modification as possible. Accordingly, Sahagian and Maus only sampled flows with a vesicularity gradient consistent with that of a simple emplacement history flow as predicted by an earlier numerical model of Sahagian *et al.*⁷. They admit nonetheless that this criterion does not “eliminate the possibil-

ity of inflation or deflation”. Clearly, much care must be taken when choosing flows to sample for this type of analysis.

The technique would benefit from more work on the controversial but critical aspect of emplacement history and from more calibration studies from modern environments. But the study by Sahagian and Maus is a good start for a promising method. □

Kathryn M. Gregory is in the Lamont-Doherty Earth Observatory of Columbia University, Box 1000, Palisades, New York 10964, USA.

DEVELOPMENTAL BIOLOGY

Insects take a homeotic test

Alfonso Martinez Arias

It is one hundred years since W. Bateson^{1,2} first highlighted what he called ‘homeotic’ variations, which we now know represent mutations in homeotic genes. In insects, these mutations can result in dramatic transformations of one body part into another: a leg instead of an antenna, or a pair of extra wings instead of small balancing organs, known as halteres. In the meantime, our ideas about the causes of these aberrations, and what they mean for the normal function of the genes affected, have evolved^{3–5}. As part of this continuing endeavour, Warren *et al.* have performed a comparative analysis of *Ultrabithorax* (*Ubx*) expression on two insects, fruitflies and butterflies (see page 458 of this issue⁶). Their results are surprising, and should prompt us to revise a widespread perception — namely, that homeotic genes operate through a sort of code that mediates ‘segment identity’³.

In two influential papers^{7,8}, E. B. Lewis summarized his work on the genetics of the bithorax complex (BX-C) of *Drosophila*, perhaps the most famous homeotic set of genes. In the second of them⁸, he distilled the essence of data and thoughts gathered during his many, distinguished years at the California Institute of Technology, and stated that:

... during the evolution of the fly, two major groups of genes must have evolved: ‘leg suppressing’ genes which removed legs from abdominal segments of millipede-like ancestors followed by ‘haltere promoting’ genes which suppressed the second pair of wings of four winged ancestors.

Lewis went on to discuss the evidence that the BX-C is involved in this process, and suggested that the differences between segments arise because particular genes of the BX-C allow the development of particular structures (today, incidentally, we know that many of his genes represent regulatory elements of a few structural genes). A segment, in this view, is a

mosaic of cell-type-specific functions of the homeotic genes. These functions are to instruct the development of various structures, for example prolegs, parts of the respiratory apparatus or sensory elements.

Further ideas as to how these functions are implemented came from A. Garcia



Flight gear — *Precis coenia* displays its four wings. Normal fruitflies have only two, the third thoracic segment producing a pair of halteres instead of wings.

Bellido, who emphasized that homeotic genes modify a basic underlying morphological pattern through the activity of downstream ‘realizator’ genes⁹. This work also established a relationship between developmental compartments and homeotic genes which was later used in a bold attempt to reduce the developmental complexity of the fly to a simple code of gene activities^{3,10}. In this view, the function of the homeotic genes is not to direct the development of particular cell types,

but to contribute to a ‘genetic address’, an abstract global property which somehow provides cells with a sense of their ‘segmental identity’³.

Mutations in the BX-C can produce that landmark in homeosis, the four-winged fly. This fly displays a spectacular extra pair of wings because its third thoracic (T3) segment has been transformed into the second (T2). The ‘mosaic’ and ‘code’ schools of thought view this animal very differently. The first has it that the extra wings result from failure to implement *Ubx* function in a particular group of cells, probably the haltere primordium, at a particular time in development; the second, that the four-winged fly results because absence of *Ubx* expression in the T3 segment makes those cells lose their ‘identity’.

Warren *et al.* have now addressed this issue by cloning the *Ubx* gene from the butterfly *Precis coenia*. In this insect, both T3 and T2 develop a pair of wings much as a four-winged fly does. The question, then, is this — is *Ubx* expressed in the T3 disc of the butterfly? Before the experiment, most punters would have put their money on the absence of expression, because that is what our knowledge of the homeotic genes in flies and their phylogenetic implications indicate. The answer, however, is that *Ubx* is expressed in the T3 dorsal disc of *P. coenia*, as it is in its fly homologue, and yet a large wing develops.

This result indicates that the presence or absence of *Ubx* is not the simple determinant of segment identity as the code model would have it^{3,10}. Expression of *Ubx* — and by extension of other homeotic genes — simply creates regulatory possibilities that will be exploited in a cell- or species-specific manner⁵. In the case of the T3 segment of the fruitfly, *Ubx* promotes halteres; in the case of the butterfly, wings.

Such cell-type-specific actions of homeotic genes are highlighted by another observation of Warren *et al.* — the appearance of rudimentary legs in the butterfly’s developing abdomen. It is clear from the work on *Drosophila* that, as postulated by Lewis^{7,8}, BX-C genes suppress leg development in the abdomen. It is therefore interesting that the development of prolegs in the abdominal seg-

1. Bateson, W. *Materials for the Study of Variation* (Macmillan, London, 1894).
2. Lewis, E. B. *Trends Genet.* **10**, 341–343 (1994).
3. Lawrence, P. A. *The Making of a Fly* (Blackwell Scientific, Oxford, 1992).
4. Akam, M. E., Dawson, I. & Tegar, G. *Development* **104** Suppl., 123–133 (1988).
5. Peifer, M., Karch, F. & Bender, W. *Genes Dev.* **1**, 891–898 (1987).
6. Warren, R. W., Nagy, L., Selegue, J., Gates, J. & Carroll, S. *Nature* **372**, 458–461 (1994).
7. Lewis, E. B. *Am. Zool.* **3**, 33–56 (1963).
8. Lewis, E. B. *Nature* **276**, 565–570 (1978).
9. Garcia-Bellido, A. *Am. Zool.* **17**, 613–629 (1977).
10. Lawrence, P. A. & Morata, G. *Cell* **35**, 595–601 (1983).
11. Castelli-Gair, J., Greig, S., Micklem, G. & Akam, M. *Development* **120**, 1983–1995 (1994).

ments of the butterfly larva is preceded by a suppression of BX-C expression, and a subsequent implementation of a thoracic programme of gene expression precisely in those cells that will give rise to prolegs. This programme includes a redeployment of homeotic gene expression in these cells.

Earlier this year, J. Castelli-Gair *et al.*¹¹ directly tested the functional importance of 'mosaic expression' of homeotic genes. Their results reinforce Lewis's ideas and show that, in *Drosophila*, the 'identity' of a segment is the result of a series of cell-type-specific intrasegmental requirements of particular homeotic genes.

This view is vindicated by the work of *Drosophila* molecular homeoboxers, who have been successfully finding cell-type-

specific targets and functions of homeotic genes. But its importance is not widely appreciated, particularly among those who study vertebrates. Perhaps unfortunately, genes do not paint stripes or draw lines, they merely encode proteins which then engage in intricate but ultimately beautiful and simple interactions. When orchestrated by cell signalling in time and space, these interactions contribute to the apparently complex patterns that developmental biologists are trying to understand. □

Alfonso Martinez Arias is a Wellcome Trust Fellow in the Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

PROTEIN-PROTEIN INTERACTION

Complex flexibility

Frances Jurnak

HUMAN growth hormone (hGH) has been a fertile system for studying signal transduction and protein-protein interactions at a high-affinity ligand-receptor interface¹⁻⁴. In the hands of Somers *et al.*, the system has delivered some fresh surprises, as described on page 478 of this issue⁵.

The hormone is a product of the pituitary, and it regulates muscle, bone and cartilage development through binding to its receptor (hGHR). It also regulates lactation by binding to a second receptor, the human prolactin receptor (hPRLR). The latter receptor, but not hGHR, is additionally activated by prolactin, another pituitary hormone. Both receptors are composed of a soluble extracellular domain, a single-pass membrane domain and a cytoplasmic domain; and both belong to the class 1 haematopoietic receptor superfamily^{6,7}, which also includes interleukins, colony stimulating factors and interferons, on account of sequence similarities in their extracellular domains.

Two years ago, de Vos, Ultsch and Kossiakoff⁸ reported the first three-dimensional structure of a ligand-receptor complex — this was hGH bound to two copies of the soluble, extracellular domain of hGHR. The results were stunning, because it was evident that the hormone formed a complex with two receptor molecules. Not only did the finding provide support for a clustering mechanism for signal transduction across membranes, but it also provided atomic details of how the same set of amino acids on the receptor could specifically recognize two different sets of amino acids on the hormone.

Now the same group, with the addition of W. Somers, has repeated the initial

success, this time with a 2.9 Å crystal structure of a 1:1 complex between hGH and the soluble extracellular domain of hPRLR. Without a second bound receptor, the complex is not active in signal transduction, but instead represents a stable intermediate in the activation pathway.

The three-dimensional structure of the hGH-hPRLR complex in itself has several interesting features. First, it explains the Zn²⁺ dependency of the hGH-hPRLR interaction. A Zn²⁺ binding site is created at the interface of the ligand-receptor interaction, and is formed by two ligands from the hormone and two from the receptor. Second, the structure confirms the amino acids identified by mutagenesis studies as the specificity determinants of the hGH-hPRLR interaction, but shows that the binding interface is much larger than was expected. Such results underscore the shortcomings of structural studies, on their own, to distinguish unambiguously between those interactions that contribute to binding and those that contribute to specificity. Finally, the complex structure has enabled the investigators to re-interpret mutagenesis data and to correct misleading conclusions formulated in the absence of structural information.

The greatest impact of the hGH-hPRLR structure lies, however, in its comparison with the hGH-(hGHR)₂ complex, allowing Somers *et al.* to consider how recognition between the same ligand surface and diverse receptor surfaces is accomplished. hPRLR is folded into two antiparallel β -domains, connected by a linker of five amino acids. The hPRLR structure is very similar to that of hGHR, despite a similarity in sequence of only 28 per cent. Not only do the β -strands in the

core superimpose well, but even the local conformations of most loops, despite a large dissimilarity in sequence, agree very well. In both receptors, six comparable loops which extend from the β -core form the binding determinant for the same set of amino acids on hGH.

The main structural differences between the two receptors lie in a rigid-body rotation of the two receptor domains with respect to hGH, as well as in the relative disposition of the amino- and carboxy-terminal domains with respect to one another. Although the rigid-body rotation alters the relative disposition of the receptor loops in the binding site, it allows a different set of interactions to be maximized. In both hGH-receptor complexes, there are nine salt bridges and hydrogen-bond pairs at the binding interface, but the composition of the interacting pairs differs. The rigid-body rotation appears to be controlled by favourable side-chain packing interactions of a single amino acid, Tyr in hGHR or Glu in hPRLR, preceding an identical sequence in the linker region of both receptors. Although it is difficult to ascertain, the rigid-body rotation of the amino- and carboxy-terminal domains with respect to one another in hPRLR may also hinder the proper formation of the second receptor binding site needed for activation. Similar rigid-body rotations of domains and the concomitant alteration in a binding surface have been observed in other systems, such as immunoglobulins⁹ and the nucleotide-controlled conformational changes observed in elongation factor-Tu¹⁰. But the unique features of the trigger sequences are not well understood or predictable and few, if any, specific principles have emerged.

Another question addressed by Somers *et al.* is the effect of receptor binding on the conformation of the hormone. hGH forms a four-helix bundle with a classical cytokine fold, and its structure is very similar in both receptor complexes. The major differences occur in poorly defined loop regions and in two small minihelices. The same regions were found to be conformationally sensitive to the local environment in a study of an uncomplexed hGH variant in two crystal forms¹¹. Only the minihelices contribute amino acids to the hPRLR binding site.

The largest change which appears to affect the receptor interaction is the unwinding of the amino-terminal turn of minihelix 1. The unwinding alters the relative orientation of minihelix 1 with respect to the receptor binding site and, consequently, causes rearrangement of the salt bridges and hydrogen bonds at the interface. Another helix unwinding occurs at the amino terminus of helix 3. This region does not interact directly with the first receptor, but may control the fate of the second receptor binding site. It is interesting to note that the unwinding of a

helix has been implicated in local conformational changes in other proteins, most notably in G proteins^{10,12}, but the process has not been systematically studied.

The final notable feature arising from the comparison of hGH complexed to two different receptors is the identification of a distinct structural motif involving the so-called WS $\times$ WS box. This sequence is common to class 1 receptors in the haematopoietic receptor superfamily, and is part of a larger, unique structural motif which covers one surface of the β -sheet in the carboxy-terminal domain of the receptor. An interleaved arrangement of aromatic and positively charged amino acids was first noted in the hGH-(hGHR)₂ complex, but its significance as a distinct motif was not fully appreciated until it appeared again in hGH-hPRL_R. The function, if any, of the WS $\times$ WS motif remains an enigma. But because the participating amino acids are conserved across the superfamily, it is likely to have more than just a structural role.

The structure of hGH-hPRL_R and its comparison with that of hGH-(hGHR)₂ provides great insight into the adaptability of the molecular recognition process. The seemingly unlimited flexibility that is possible complicates the search for specific principles which might underlie protein-protein interactions. Two types of movement observed in hGH-receptor complexes — rigid-body rotation of domains and helix winding/unwinding — are becoming more evident, though the rules involving the optimal side-chain interactions that determine specificity are more elusive with so few examples studied. The problem is similar to recognition specificity between DNA and proteins, an area which has seen major advances in the past 15 years. But with 20 amino acids possible on each side of the paired interactions, the problem is much greater for protein-protein interactions, meaning that many more examples of protein complexes will need to be determined in atomic detail before any specificity principles emerge. □

Frances Jurnak is in the Department of Biochemistry, University of California, Riverside, California 92521, USA.

Trials in bad taste

Peter D. Moore

ALTHOUGH many herbaceous plants, particularly those with growing points close to the ground, cope well with persistent grazing from mammalian herbivores, woody plants are generally less capable of withstanding the attention of browsers. One of the most effective defence mechanisms is the employment of secondary metabolites as grazing deterrents¹. Such chemical defence is widely recorded in plants, but little is currently known of

forests most of the canopy herbivory results from the activity of invertebrates, but juvenile trees and shrubs, together with lower branches of mature trees, may also be subjected to mammalian browsing, and mammals can also be deterred by unpalatable compounds.

Biogeographical studies of such herbivore deterrence has indicated that there is considerable variation of palatability of juvenile individuals among closely related

tree species in different parts of the warm temperate and boreal zones. Could such variation be due to a long exposure to browsing in the past history of these plants? Are those areas that are known to have been subjected to intense herbivory over a long period of time the sites where the greatest levels of grazing resistance are found? Or do biogeographical patterns reflect current browsing activities or densities of herbivorous mammals? Bryant and co-workers have now conducted more detailed surveys of North America

The bitter about to be bit — a snowshoe hare pursued by a bobcat, *Felis rufus*.

geographical variation in the intensity of such mechanisms, or the possible causes of any biogeographic patterns. In a recent issue of *Oikos*, Bryant, Swihart, Reichardt and Newton^{2,3} describe feeding experiments with snowshoe hares by which they show that there is strong latitudinal and longitudinal variation in the palatability of the juvenile stages of aspen and birch trees. They account for the trends in terms of past and present levels of herbivory, which together have selected for a greater intensity of chemical defence in northern and northwestern areas of North America.

Woody plants are susceptible to the removal of their shoots because the growing points (buds) are borne upon them and their loss represents an immediate energetic depletion coupled with the destruction of much of the plant's recovery potential. This is in contrast to certain herbaceous plants, especially grasses, that grow from the base and are resilient in recovery terms even when they experience regular energy losses as a result of shoot removal by grazing or mowing. One evolutionary response of woody plants to persistent herbivory is the accumulation of secondary products within twigs and leaves that consequently become unpalatable or even toxic to the grazers, an example being the tannins found in the leaves of oaks. In temperate and boreal

to determine how extensive is the variation in unpalatability and whether there is an underlying latitudinal or longitudinal pattern.

Using aspen (*Populus*) and birch (*Betula*) species from Alaska and Maine, they carried out 'cafeteria-style' feeding trials, offering mature and juvenile shoots to both free-ranging and captive snowshoe hares in the two areas under study². In all trials the hares ate fewer of the juvenile shoots of the Alaskan aspen and birch (*P. tremuloides* and *B. resinifera* respectively) than of those from Maine (*P. grandidentata* and *B. papyrifera*, *B. allegheniensis* and *B. lenta*). There was little variation in preference for adult shoots. When seeking an explanation for this longitudinal pattern, the authors consider the possibility that the megafaunal herbivores of the last glaciation could be responsible. But if the Pleistocene grazing pressures were the stimulus behind grazing resistance, then the eastern species should be more unpalatable as these regions would have been subjected to glacial browsing by mastodons and ground sloths in their refugia south of the Laurentide Ice Sheet⁴. Beringian mammoths, horses and bison were grazers rather than browsers with few, if any, trees being available to them. Because the northwestern species currently have the less palatable juvenile shoots, this hypothesis must be rejected.

Oxford Scientific Films

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

- Chawla, R. K., Parks, J. S. & Rudman, D. A. *Rev. Physiol.* **47**, 483–499 (1983).
- Cunningham, B. C. & Wells, J. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3407–3411 (1991).
- Cunningham, B. C., Bass, S., Fuh, G. & Wells, J. A. *Science* **250**, 1709–1712 (1990).
- Sprang, S. R. & Bazan, J. F. *Curr. Opin. struct. Biol.* **3**, 815–827 (1993).
- Somers, W., Ultsch, M., De Vos, A. M. & Kossiakoff, A. A. *Nature* **372**, 478–481 (1994).
- Bazan, J. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7934–7938 (1990).
- Cosman, D. *Cytokine* **5**, 95–106 (1993).
- De Vos, A. M., Ultsch, M. H. & Kossiakoff, A. A. *Science* **255**, 306–312 (1992).
- Brünger, A. T. *Immunomethods* **3**, 180–190 (1993).
- Berchtold, H. *et al.* *Nature* **365**, 126–132 (1993).
- Ultsch, M. H., Somers, W., Kossiakoff, A. A. & De Vos, A. M. *J. molec. Biol.* **236**, 286–299 (1994).
- Jurnak, F., Heffron, S. & Bergmann, E. *Cell* **60**, 525–528 (1990).

If the periglacial explanation is unacceptable, the alternative is that the pattern is caused by post-glacial and current browsing intensity in Alaska, mainly by post-fire herbivores such as moose and snowshoe hare. The hare is known to undergo population cycles in Alaska in which peak densities can be ten times those of the highest values recorded in Maine, so periodic high-intensity grazing may be experienced by the Alaskan trees. The fire cycles in this area are known to go back well into the Holocene⁵, so these acute grazing episodes may also be of long standing.

This longitudinal study has been supplemented by a latitudinal survey conducted in the eastern United States, from Maryland north through Connecticut into Maine, using juvenile and mature twigs of the two aspens and the three eastern birches. Again, captive and free-ranging snowshoe hares were used as consumers. The outcome was a clear avoidance of juvenile shoots of northern species (*B. papyrifera* and *P. tremuloides*), variable response in the intermediate-latitude yellow birch (*B. allegheniensis*), but no evident preference for juvenile material in the southern species (*B. lenta* and *P. grandidentata*). Samples of a single species taken from different latitudes also produced similar results, with the hares eating less of the juvenile material from the north, but failing to discriminate between mature fodder from different latitudes.

Clearly, there is a latitudinal influence on the palatability of juvenile shoots, but one has to take into consideration the direct effects of climate on twig and leaf chemistry. Perhaps there is a drought- or frost-hardiness associated with some secondary products which also happen to deter grazers. If this were so, however, according to the arguments of Bryant and co-workers, the effect would show up in the mature as well as the juvenile samples. Indeed, one might expect a greater exposure to drought and frost higher in the tree canopies, so a greater accumulation of protective materials there. The fact that it is juvenile shoots that display the grazing deterrence adds weight to the argument that the latitudinal gradient is herbivore-induced and also that the herbivores are found on the ground rather than in the canopy.

Many interesting questions are generated by these studies. Are mammalian herbivores more important as selective forces in tree evolution in the boreal zone than they are in warmer temperate regions? Are the interspecific differences

between the trees described here a reflection of particular pressures exerted by browsers at the time of their evolutionary divergence? And how will climate change (should it come about) influence the balance of tree and herbivore in the northern realms? Will ill-equipped, unprotected

southern birches march northwards to their inexorable fate beneath the incisors of ravenous boreal hares? □

Peter D. Moore is in the Division of Life Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

GALAXY EVOLUTION

Chronology of the skies

Ralph E. Pudritz

ONE of the great riddles in astrophysics is how galaxies and their stars are born. A traditional approach is to use the relative abundances of elements and their isotopes within stars and meteorites as chronometers for these processes, and, as results presented at a recent workshop* indicated, there is much new progress in the field. Among the advances is evidence that our Solar System formed in a durable, long-lived massive star complex that had been lashed by bursts of supernova explosions. It may have had many of the characteristics of the nearby Orion molecular cloud.

Elements beyond iron are produced by the capture of neutrons. Neutron capture on rapid (r-process) or slow (s-process) timescales with respect to the time for β -decay occurs in supernova explosions or

stellar ejecta and supernovae have specific metallic and isotopic 'fingerprints' that are impressed upon subsequent star formation in the interstellar medium.

Now consider the chronology of the formation of the Milky Way. Here one can use the globular clusters that inhabit the halo and central bulge of our Galaxy and that are the oldest known stellar subsystems (the most ancient thought to be 16 ± 2 billion years old¹). The protoglobular parental gas for these first stars had a heavy-element abundance less than a hundredth that of the Sun. Because the very first period of chemical enrichment is due to the short-lived massive stars that end up as type II supernova, one expects that the stars in the most metal-poor globular clusters have compositions dominated by r-process elements. The s-process abund-

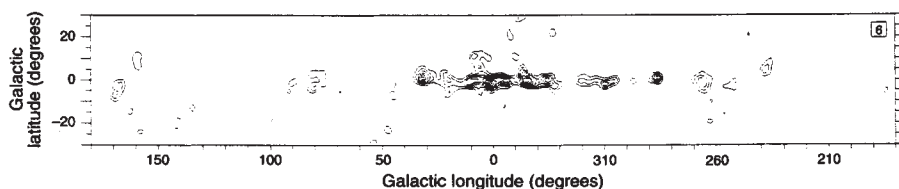


FIG. 1 The Milky Way in the light of the 1.809-MeV line from the decay of ^{26}Al , obtained with the COMPTEL telescope (adapted from ref. 6).

in stellar interiors, respectively. Supernovae (or at least type II supernovae) from massive stars between 10 and 30 solar masses ($M_{\odot}$) dominate the metallicity of the Galactic interstellar medium out of which the Sun and meteorites in our Solar System formed, and can be traced through the r-process elements. Intermediate mass stars ($1-10 M_{\odot}$), including stars on the asymptotic giant branch of evolution, shed their element-enriched outer layers as winds and planetary nebulae. They contribute predominantly s-process elements to the interstellar medium leaving white dwarfs behind. Finally, white dwarfs in old binary stellar systems eventually accrete gas from their companions, exceed their stability limits, and go supernova (type Ia). These are primarily responsible for the iron-peak elements in the interstellar medium. Thus

ances ought to appear later in the chronology when the longer-lived lower-mass stars go into their asymptotic giant branch and supernova type Ia phases².

These ideas appear to fit the general observational trends (J. Truran, Univ. Chicago). However, translating this into a more detailed chronology of the galaxy's formation is tricky because the complicated process of mixing the metal-enriched supernova ejecta with the bulk of the metal-poor gas takes a considerable amount of time (D. Clayton, Clemson Univ.). If instantaneous mixing is assumed, the initial collapse and burst of star formation in the Milky Way may have been over in a billion years or less. But realistic mixing times imply that the Galaxy probably formed in a more prolonged manner. An attractive interpretation of the large cluster-to-cluster scatter in metallicities is that the Galaxy built up more slowly by the accumulation of smaller clusters containing fragments

1. Bryant, J. P. et al. *A. Rev. ecol. Syst.* **22**, 431-446 (1991).
2. Bryant, J. P., Swihart, R. K., Reichardt, P. B. & Newton, L. *Oikos* **70**, 385-395 (1994).
3. Swihart, R. K., Bryant, J. P. & Newton, L. *Oikos* **70**, 427-434 (1994).
4. Pielou, E. C. *After the Ice Age: The Return of Life to Glaciated North America* (Univ. Chicago Press, 1991).
5. Viereck, L. A. *Quat. Res.* **3**, 365-395 (1973).

*Workshop on the Evolution of Star Forming Regions, Harvard-Smithsonian Center for Astrophysics, Boston, USA, 24-26 October 1994.

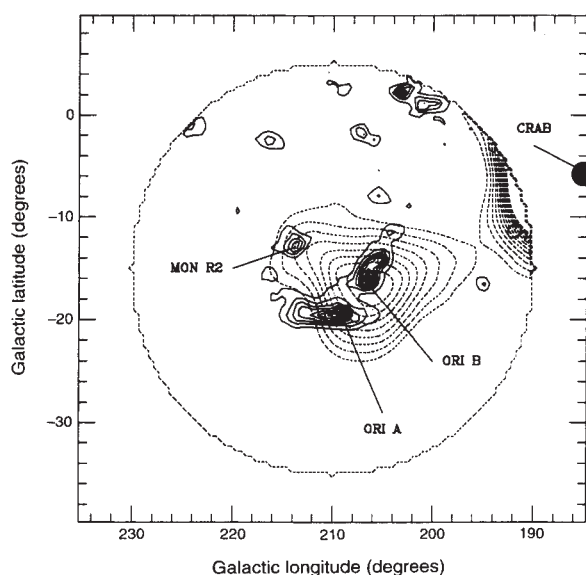


FIG. 2 COMPTEL map of the Orion region in the 3–7-MeV range (dotted contours) identified with the 4.44- and 6.13-MeV de-excitation lines of carbon and oxygen nuclei. The Columbia University carbon monoxide observations are superimposed for comparison (from ref. 7).

each with its own star formation history³. Good candidates for these parental fragments are super-massive self-gravitating clouds whose physical properties resemble those of current molecular clouds such as Orion⁴.

Most encouragingly, the processes of enrichment and mixing can now be observed through the direct imaging of nuclear de-excitation γ -ray lines by instruments such as the COMPTEL telescope on board the Compton Gamma Ray Observatory (H. Bloemen, SRON-Utrecht). One example is the radioactive decay of ^{26}Al by positron emission into an excited state of ^{26}Mg . This process has a half-life of about a million years and the resultant 1.809-MeV line has now been mapped along the Milky Way (Fig. 1). Although the precise origin of the ^{26}Al is not established yet, the extended and structured appearance of the Milky Way observed at 1.809 MeV suggests that regions where massive stars form play a dominant role through their production of supernova type II ejecta and stellar (Wolf-Rayet) winds. Another important result is the possible COMPTEL detection of nuclear interaction lines of carbon and oxygen nuclei from the Orion complex at 4.44 and 6.13 MeV (Fig. 2).

The patchy nature of the COMPTEL ^{26}Al map suggests that the amount of that isotope ($1.5 M_{\odot}$) spread throughout the interstellar medium is a product of nucleosynthesis in massive stars. But the inferred average ratio of $^{26}\text{Al}/^{27}\text{Al}$ in the interstellar medium is still an order of magnitude smaller than the large ratio, 5×10^{-5} , found in meteorites that formed in the early solar disk. Clayton⁵ argued that the low-energy cosmic rays responsi-

ble for the 4.43-MeV γ -rays detected by COMPTEL from the Orion clouds are also responsible for the ^{26}Al in molecular cloud cores, as supernova models have difficulty in accounting for this high ratio through self-enrichment. Another possibility is that particles from the enriched hot medium are bombarding the ambient dense gas, after having been accelerated by colliding stellar winds and multiple supernova explosions (Bloemen).

The process of star formation can be studied through the isotopic abundances within meteorites that locked in information about conditions prevalent at the birth of our Solar System. The inferred initial abundances of most of the shorter-lived radionuclides such as ^{60}Fe , ^{53}Mn and ^{107}Pd (half-life of 6.5 million years) is consistent with a burst of self-enrichment by type II supernovae. The Solar System may therefore have formed in the compressed wall of metal-enriched gas that abuts a region of vigorous O and B star formation as is occurring in the Orion molecular cloud.

The presence of long-lived isotopes such as ^{129}I (half-life of 16 million years) is also very important (C. L. Harper Jr, Harvard Univ.). Thus, whereas the short-timescale data indicate an episode of self-enrichment of the parental cloud, the inference from the data on iodine with its longer decay time is that the parental cloud underwent a long period of isolation from type II bursts. This could occur naturally for parental gas that was overtaken by a spiral arm of the Galaxy. The denser regions of high-mass star formation and type II supernovae in the spiral arms would produce spikes in the ^{129}I record followed by long quiet spells while the cloud was between arms (C. L. Harper, manuscript submitted).

The Orion region may not only be presenting us with a picture of our own violent origins. It may also be a laboratory for star-formation processes that occurred when our Galaxy was assembling. □

Ralph E. Pudritz is in the Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada.

1. Bolte, M. *ASP Conf. Proc.* **48**, 60 (1993).
2. Wheeler, J. C., Sneden, C. & Truran, J. W. Jr. *A. Rev. Astr. Astrophys.* **27**, 279 (1989).
3. Searle, L. & Zinn, R. *Astrophys. J.* **225**, 357 (1978).
4. Harris, W. E. & Pudritz, R. E. *Astrophys. J.* **429**, 177 (1994).
5. Clayton, D. D. *Nature* **368**, 222 (1994).
6. Diehl, R. et al. *Astr. Astrophys.* (in the press).
7. Bloemen, H. et al. *Astr. Astrophys.* **281**, L5 (1994).

Navigation grid

LIKE most developed regions, Europe is covered with a dense power grid. It carries gigawatts of 50-Hz alternating current through lines often hundreds of kilometres long. The resulting a.c. magnetic field must extend around the lines for some function of this characteristic distance. It must blanket the whole continent, and even reach up into space. Daedalus now has a use for it. He sees it as a huge distributed European radar aerial.

Unlike conventional radar, it would work not by radiation but by induction. The grid acts as the primary winding of a vast transformer. Any metal object within its field — such as an aircraft at a height of a few kilometres — acts as a secondary winding: a current is induced in it. This in turn perturbs the field. By measuring the field at many points, it should be possible to detect this perturbation and deduce its source. To complete the radar, therefore, a set of sensing wires is needed. The grid pylons themselves could easily carry an extra wire or so, and many other existing conductors could be pressed into service. Telephone lines, data links, wire fences, railway signal cables, motorway dividing strips and so on, all could pick up their local 50-Hz field and monitor its direction, amplitude and phase. These measurements could be relayed back to a central computer, and compared with the currents flowing at that moment within the grid itself. In the absence of air traffic, they would define the background field of the continent. If aircraft now flew over the system, they would perturb the field. Subtracting the known unperturbed field would reveal these perturbations, and locate the singularities from which they emanated: the aircraft.

For air-traffic control, this system has many advantages. Europe could replace its many control centres with one big network. It could be extended into any other region which had a 50-Hz grid kept in phase with the European one. Aircraft using the system could enhance their visibility by carrying tuned coil-and-capacitor 50-Hz resonators, but even Stealth bombers should show up clearly enough. Indeed, for air defence the continental hum field would be ideal. It is everywhere all the time, it is too big and widely distributed to be knocked out, it gives away no information of its own, and nothing can fly under or over it. The only snag is that it would detect surface traffic as well. This could overload the computation. But, says Daedalus, the low speed and defined tracks of surface traffic make it easy to recognize and subtract from the perturbation analysis at an early stage.

David Jones

DNA damage and melanogenesis

SIR — The action spectrum and histological manifestations of tanning have been well studied¹, but the mechanism by which ultraviolet exposure stimulates melanin production is unknown. Ultraviolet irradiation directly stimulates melanogenesis in pigment cells², and T4 endonuclease V (T4N5), a DNA repair enzyme, enhances UV-induced melanogenesis in cultured melanocytes³. Because T4N5 is known to augment the rate-limiting step in excision repair of UV-induced photoproducts, we hypothesized that DNA repair is the initial signal for increased melanin synthesis following ultraviolet exposure and that increased excision of DNA fragments containing these lesions is responsible for the enhanced tanning response of irradiated endonuclease-treated pigment cells.

In both prokaryotes and eukaryotes,

excision repair generates a fragment of single-stranded DNA containing the photodamaged nucleotides^{4,5}. We hypothesized that the basal tanning response, as well as its enhancement by T4N5 treatment, is initiated by excision of the DNA fragments containing the photoproduct, usually formed at pyrimidine dinucleotides⁶. We therefore analysed the effect of transfection with exogenous fragmented, phospholipid-encapsulated salmon sperm DNA (a gift of Dr Gerard Redziniak) on Cloudman S91 murine melanoma cells (S91 cells). The melanin content of the DNA-treated cells increased, but the magnitude of the response was variable.

Because ultraviolet irradiation of DNA preferentially induces a dipyrimidine photoproduct, we assayed dipyrimidine dithymidylic acid (pTpT) for its effect on

melanin synthesis in mouse melanoma cells. We used deoxyadenylic acid dinucleotide (pdApdA), a purine dinucleotide not significantly dimerized by ultraviolet irradiation, as a control. After 4 days, differences in melanin content were evident on visual inspection of the cell pellets (Fig. 1a). Cells treated with pTpT had an approximately sevenfold higher melanin content than control cells, whereas pdApdA stimulated only a minimal increase in melanin content (Fig. 1b). pTpT also increased melanogenesis in cultured human epidermal melanocytes (data not shown).

To explore the mechanism of pTpT-induced pigmentation, we maintained paired cultures of S91 cells for 4 days in pTpT, pdApdA or diluent alone and used northern blots to determine the expression of the messenger RNA encoding tyrosinase, the rate-limiting enzyme, in melanogenesis. Cells treated with pTpT showed a two- to threefold increase in tyrosinase mRNA level over control cells at 72 and 96 hours (Fig. 1c), a time course consistent with a causal role in the resulting melanogenic response of the treated cells. Cells treated with pdApdA or diluent showed minimal or no increase, respectively, of either tyrosinase mRNA or melanin per cell.

The tanning response is characterized clinically by a brown colour in ultraviolet-exposed skin and histologically by an increase in epidermal melanin content¹. To determine if pTpT produces a tanning response in intact skin, we applied dissolved dinucleotide to shaved guinea-pig skin twice daily for 5 days. Increased pigmentation in treated skin sites was first observed after 10–14 days and persisted for at least 60 days. Applications of 100 and 300 μ M pTpT in either vehicle increased pigmentation compared with untreated and vehicle-treated sites (Fig. 2a). Histological sections demonstrated an increase in melanin throughout the epidermis of the pTpT-treated sites, particularly in the basal layer, compared with untreated and vehicle-treated skin (Fig. 2b–d), identical to the appearance of ultraviolet-irradiated, tanned skin. Application of pdApdA did not produce a tanning response.

Although the molecular signal initiating the ultraviolet response is unknown, our data suggest that one signal for melanogenesis following ultraviolet irradi-

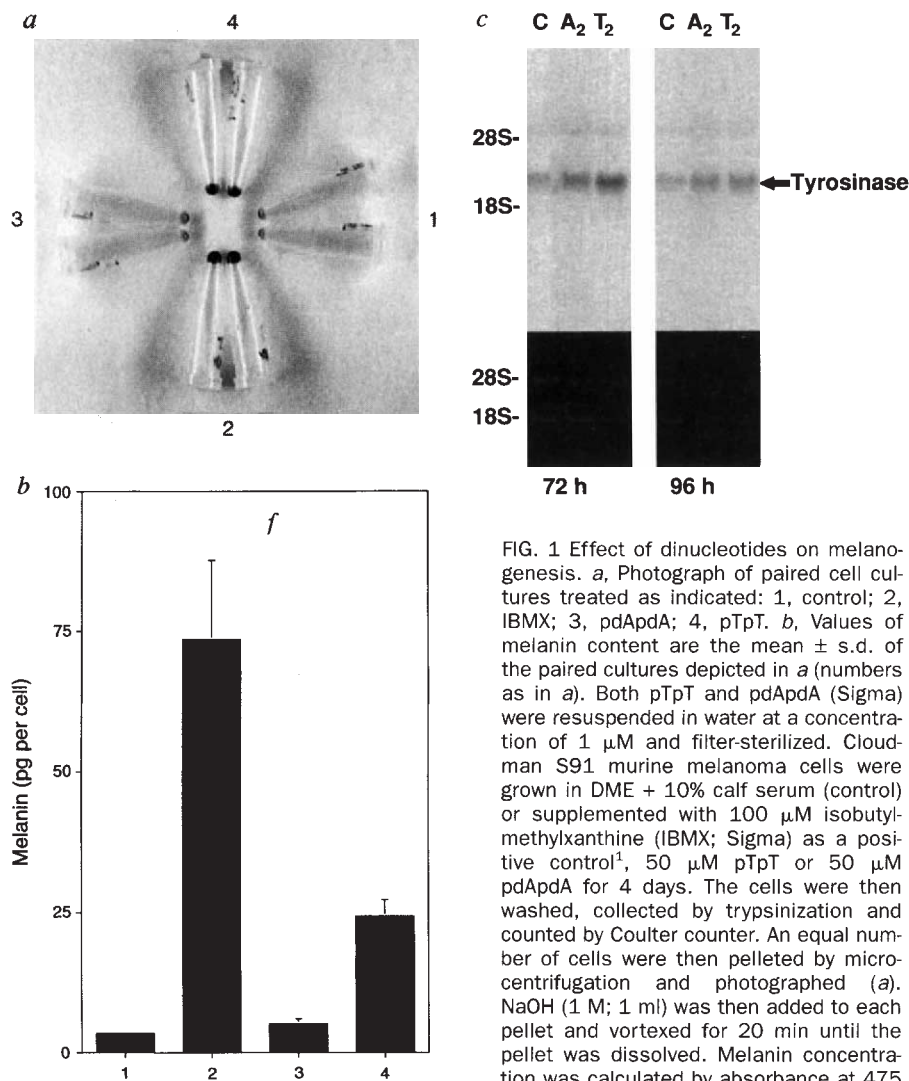


FIG. 1 Effect of dinucleotides on melanogenesis. *a*, Photograph of paired cell cultures treated as indicated: 1, control; 2, IBMX; 3, pdApdA; 4, pTpT. *b*, Values of melanin content are the mean $\pm$ s.d. of the paired cultures depicted in *a* (numbers as in *a*). Both pTpT and pdApdA (Sigma) were resuspended in water at a concentration of 1 μ M and filter-sterilized. Cloudman S91 murine melanoma cells were grown in DME + 10% calf serum (control) or supplemented with 100 μ M isobutylmethylxanthine (IBMX; Sigma) as a positive control¹, 50 μ M pTpT or 50 μ M pdApdA for 4 days. The cells were then washed, collected by trypsinization and counted by Coulter counter. An equal number of cells were then pelleted by microcentrifugation and photographed (*a*). NaOH (1 M; 1 ml) was then added to each pellet and vortexed for 20 min until the pellet was dissolved. Melanin concentration was calculated by absorbance at 475 nm compared with a standard curve of

melanin (Sigma) in 1 M NaOH as described³. *c*, In a separate experiment, paired S91 cultures were supplemented with 100 μ M pTpT (T₂), 100 μ M pdApdA (A₂) or diluent (control, C). Total cellular RNA was collected at intervals for 96 hours for northern blot analysis (10 g total RNA each sample) using a mouse tyrosinase cDNA (ATCC, cat # 63164).

1. Kochevar, I., Pathak, M. & Parrish, J. in *Dermatology in General Medicine* Vol. 1 (eds Fitzpatrick, T. et al.) Ch. 131 (McGraw-Hill, New York, 1993).
2. Friedmann, P. & Gilchrist, B. A. *J. cell. Physiol.* **133**, 88–94 (1987).
3. Gilchrist, B. A., Zhai, S., Eller, M. S., Yarosh, D. B. & Yaar, M. *J. invest. Derm.* **101**, 666–672 (1993).
4. Grossman, L., Caron, P. R., Mazur, S. J. & Oh, E. Y. *FASEB J.* **2**, 2696–2701 (1988).
5. Tateishi, S., Hori, N., Ohtsuka, E. & Yamaizumi, M. *Biochemistry* **32**, 1541–1547 (1992).
6. Mitchell, D. L. & Karentz, D. in *Environmental UV Photobiology* (eds Young, A. et al.) 345–377 (Plenum, New York, 1993).

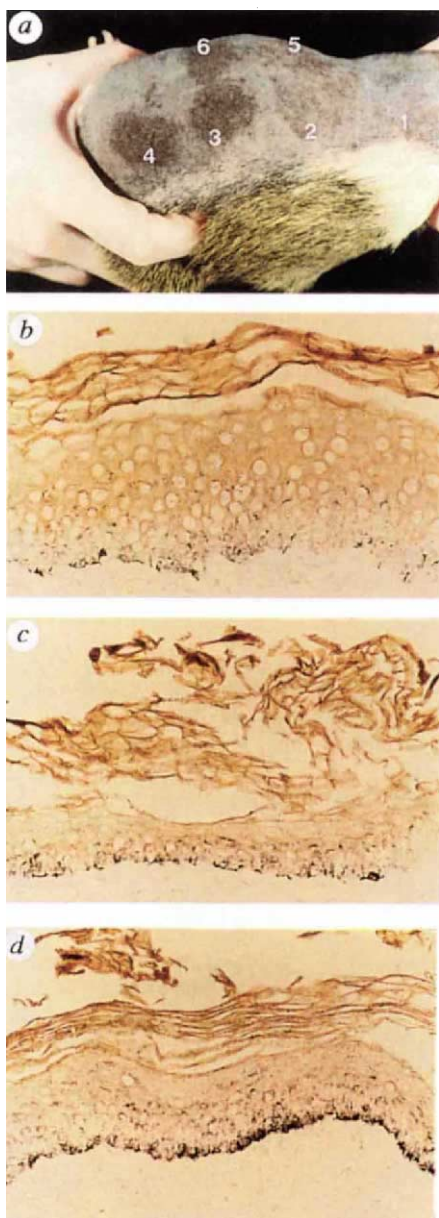


FIG. 2 Stimulation of tanning in guinea-pig skin by thymidine dinucleotides. *a*, Photograph of treated animal showing treatment sites: (1) propylene glycol (PG) alone; (2) PG/DMSO alone; (3) 100 μ M pTpT in PG; (4) 100 μ M pTpT in PG/DMSO; (5) 300 μ M pTpT in PG; (6) 300 μ M pTpT in PG/DMSO. *b*–*d*, Micrographs ($\times 18$ original magnification) showing untreated skin (*b*), vehicle-treated skin (site 1, PG; *c*) and dinucleotide-treated skin (site 3, 100 μ M pTpT in PG; *d*). American short-hair pigmented guinea pigs were obtained from Kuiper Rabbit Ranch, Chicago. Five animals were prepared by shaving and the stubble removed with a depilatory (Nair); pTpT was dissolved in either PG or a 75%/25% PG/DMSO mixture at final concentrations of 100 and 300 μ M. Applications of 25 ml were made at sites 1–6 twice daily for 5 days. The animals were shaved, destubbed and observed weekly. Twenty-four days after the final application, the animals were photographed (*a*) and biopsied. Four 4-mm punch biopsy specimens were fixed overnight in 10% formalin, processed as 4- μ m cross-sections, and stained with Fontana–Masson to highlight melanin.

ation is the excision of DNA photo-products and/or accumulation of these DNA fragments. We postulate that treatment with the pyrimidine dinucleotide pTpT mimics this signal. Furthermore, independent of this hypothesis, topical application of pTpT or other dinucleotides or DNA fragments may provide a protective tan for human skin without prior exposure to damaging ultraviolet irradiation.

Mark S. Eller

Mina Yaar

Barbara A. Gilchrest

Boston University

School of Medicine,

Boston,

Massachusetts 02118, USA

Rapid motion of liquid drops

SIR — It has been known for many years that surface tension gradients can cause liquids to flow against the force of gravity. The 'tears' that form in a wine glass are a classical example. In the meniscus where the wine meets the glass, evaporation of the alcohol leads to a surface tension gradient that draws liquid up the side of the glass. Gravity eventually wins and the wine returns to the glass in the form of tears¹. Although the basic physics underlying such phenomena was explained last century by Marangoni, there has recently been a revival of interest in "Marangoni effects"². Several groups have studied the motion of drops in the presence of thermal gradients and on chemically heterogeneous surfaces^{3,4}, and have developed theoretical models for the observed motion^{5,6}. Here we describe experiments showing that the energy released by chemical reactions with the surface can cause drops of a liquid to move rapidly across a surface, along a fibre and even up a vertical tube.

The figure shows a photograph of a drop of decane containing a fluorinated fatty acid, $\text{CF}_3(\text{CF}_2)_6\text{CO}_2\text{H}$ (PFOA), flowing up an ordinary glass microscope slide. Three frames separated by 0.45 s have been superimposed: the average speed between the first two exposures is 2 cm s^{-1} . On a level surface, the drops move twice as fast. The high speeds attained by the drops are initially surprising, but they are, in fact, in excellent agreement with theoretical predictions^{5–7}. The drops can also flow large distances. Because a pre-existing surface tension gradient is not required, the distance the drop can travel is limited only by exhaustion of the reactive ingredient dissolved in the liquid.

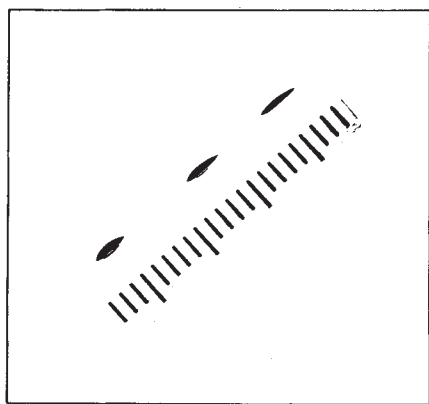
The key to this unusual behaviour, which we call 'reactive flow', is the chemistry occurring at the surface of the glass.

PFOA adsorbs from decane to form an oriented monolayer attached to the glass through the carboxylic acid group. The outer surface of the monolayer consists of oil-repelling CF_2 and CF_3 groups. Decane spreads over the surface of clean glass but beads up on a fluorinated surface, which has a much lower surface free energy. At the advancing (right-hand) edge of the drop in the figure, the glass is clean, so the drop spreads. Underneath the drop the PFOA in solution reacts with the glass. By the time the left-hand edge of the drop reaches any point on the surface, the glass has been coated with a monolayer PFOA, so the receding edge of the drop retracts, leaving monolayer-coated glass behind⁸.

It is easy to see how drop motion can be sustained by the chemical reaction, but how does one initiate motion? On a flat sheet of glass, a drop of PFOA in decane simply spreads over the clean glass in all directions, reacts, and then retracts again. If, however, we introduce asymmetry into the surface by covering part of the glass with an oil-repelling film (in this case, a fluorinated alkylsilane), then the drop is constrained to spread in only one direction, which we choose here to be 'uphill'.

Reactive flow is not restricted to PFOA on glass, but is remarkably widespread. It was first clearly identified for solutions containing alkanethiols placed on the surface of gold⁹. We have studied drop motion with fatty acids (both fluorinated and non-fluorinated) on silver and aluminium, with alkanethiols on silver and gold, and with both alkyltrichlorosilanes and fluorinated fatty acids on silicon and glass¹⁰. Reactive flow can also occur on organic surfaces: drops of PFOA in decane moved across a gold surface covered with a pre-adsorbed monolayer of $\text{HO}(\text{CH}_2)_{11}\text{SH}$. The carboxylic acid head-groups of the PFOA presumably form hydrogen bonds to the hydroxyl groups of the monolayer. In each case, reactive flow occurred in hydrocarbon solvents with suitable surface tensions. For carboxylic acids and alkanethiols on metal surfaces, some polar solvents could also be used. Acetonitrile, in particular, has a very low viscosity and consequently produced some of the fastest moving droplets.

1. Adamson, A. W. *Physical Chemistry of Surfaces*, 5th edn. (Wiley, New York, 1990), 117.
2. de Gennes, P. G. *Rev. mod. Phys.* **57**, 827–863 (1985).
3. Chaudhry, M. K. & Whitesides, G. M. *Science* **256**, 1539–1541 (1992).
4. Ondarçuhu, T. & Raphael, E. C. *rebd. Séanc. Acad. Sci. Ser. II* **314**, 453–456 (1992).
5. Brochard, F. *Langmuir* **5**, 432–438 (1989).
6. Brzoska, J. B. Brochard-Wyart, F. & Rondelez, F. *Langmuir* **9**, 2220–2224 (1993).
7. Montgomerie, R. R. thesis Oxford Univ. (1994).
8. Bigelow, W. C., Pickett, D. L. & Zisman, W. A. *J. Colloid Sci.* **1**, 513–538 (1946).
9. Bain, C. D. & Whitesides, G. M. *Langmuir* **5**, 1370–1378 (1989).
10. Burnett-Hall, G. D. thesis, Oxford Univ., (1993).



Photograph of a 2- μ l drop of decane containing 1.5 mM pentadecafluorooctanoic acid (PFOA) flowing up a glass microscope slide inclined at an angle of 43° from the horizontal. Three photographs taken at 0.45-s intervals have been superimposed in the same image: the first photograph is on the left. A millimetre-scale is shown beneath the drops. These images show both the drop and its reflection in the glass: the upper part of the lens is the real drop. The glass surface itself is not resolved in the photographs.

METHODS. To prepare the surface, a glass microscope slide is first coated with a monolayer of $\text{CF}_3(\text{CF}_2)_5(\text{CH}_2)_2\text{SiCl}_3$, on which decane does not spread. The slide is then etched through a mask in an oxygen plasma to create a pattern of tracks of clean glass each 2 mm wide and 32 mm long. A 2- μ l drop is formed at the end of a PTFE-coated needle and placed on one end of a strip to restrict flow of the drop to one direction only. The moving drop intercepts a HeNe laser beam, which triggers an automatic camera to take a series of frames. The timing between each frame is measured electronically so that the speed of the drop can be computed. PFOA in decane yields drop speeds that lie within a convenient range for our camera.

There are two principal requirements for reactive flow to occur. First, the surface chemical reaction must be accompanied by a change in interfacial free energy so that the equilibrium advancing contact angle of the pure liquid on the clean surface is lower than the receding contact angle on the monolayer-coated surface. Second, the viscosity of the solvent and the concentration of the reactant must be chosen so that there is a balance between the rate of spreading and the kinetics of the chemical reaction. Given these relatively simple restrictions, it is likely that further investigations will uncover reactive flow with many other systems, including perhaps aqueous solutions.

Colin D. Bain

Graham D. Burnett-Hall

Richard R. Montgomerie

Physical and Theoretical

Chemistry Laboratory,

University of Oxford,

South Parks Road,

Oxford OX1 3QZ,

UK

SNARE motif and neurotoxins

SIR — Tetanus and botulinum neurotoxin types A–G cause the paralytic syndromes of tetanus and botulism, respectively, by blocking neurotransmitter release¹. These toxins are zinc endopeptidases acting in the neuronal cytosol: tetanus and botulinum B, D, F and G attack specifically VAMP (also called synaptobrevin), a protein of the synaptic vesicles; botulinum A and C cleave SNAP-25; and botulinum E acts on syntaxin, both proteins of the presynaptic membrane^{1,2}. This demonstrates that VAMP, SNAP-25 and syntaxin are central in neuroexocytosis^{2,3}. VAMP has been suggested to act as vesicular receptor (v-SNARE) and SNAP-25 and syntaxin as target membrane receptor (t-SNARE) in vesicle docking and fusion within the cell^{3–6}.

Each botulinum toxin cleaves a different peptide bond within the target molecule and fails to hydrolyse short peptides encompassing the cleaved bond^{1,7–10}. Moreover, these toxins hydrolyse only one out of several identical peptide bonds present in the target protein¹. These features indicate that the cleaved region is not the sole determinant of the specific recognition of tetanus and botulinum toxins. Sequence analysis of clostridial neurotoxins suggests that they evolved from a common ancestor gene and have conserved the active-site zinc-binding motif as well

as the amino-terminal part, putatively involved in substrate recognition. Conversely, the complementary substrate-binding region must be conserved among VAMP, SNAP-25 and syntaxin. Figure 1 shows that the three putative SNARE proteins share a common motif. There are two copies of the motif in syntaxin (X1 and X2), two copies in VAMP (V1 and V2) and four copies in SNAP-25 (S1, S2, S3 and S4). The motif is predicted to adopt an α -helical configuration: Fig. 1 shows a concentration of negative charges on one-third of the surface and of hydrophobic residues on the contiguous third of the cylinder.

To evaluate the possible relevance of these conserved segments in neurotoxin recognition, corresponding peptides were synthesized (sequences and symbols in Fig. 1b) and tested as inhibitors of neurotoxin activity *in vitro* and *in vivo* and as immunogens.

Figure 2a shows that VAMP-derived peptide V2 not only inhibits botulinum-B cleavage of VAMP, but also the botulinum-A cleavage of SNAP-25 and the botulinum-C proteolysis of syntaxin. Conversely, SNAP-25-derived peptide S3 inhibits A-induced cleavage of SNAP-25 as well as B-cleavage of VAMP and C-proteolysis of syntaxin. Tetanus toxin behaves as botulinum B; S1, S2 and V1 give similar results; X1 and S4 are water-insoluble; and scrambled peptides have no effect.

Figure 2b shows the pattern of peptide inhibition of neurotoxin action in injected neurons of the buccal ganglion of *Aplysia*

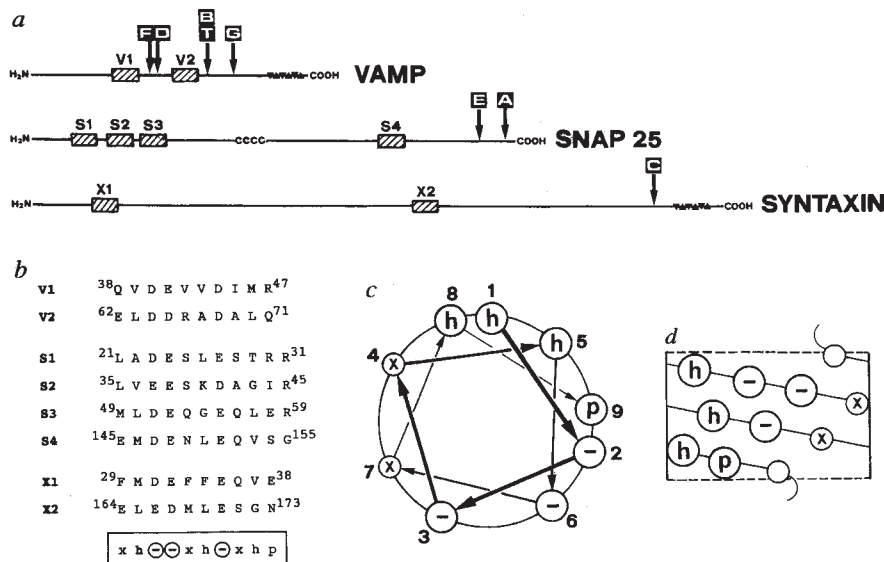


FIG. 1 A common motif is present in VAMP/synaptobrevin, SNAP-25 and syntaxin. **a**, **b**, The motif is composed of nine residues and consists of a conserved hydrophobic residue (h), two carboxylates (–), any residue (x), a hydrophobic residue conserved in all but two segments (h), a carboxylate (–), x, h and a polar residue (p). This motif is present in two copies in VAMP (V1 and V2), two copies in syntaxin (X1 and X2) and four copies in SNAP-25 (S1, S2, S3 and S4). Capital letters, site of proteolytic cleavage of tetanus toxin (T) and of the seven botulinum neurotoxin types from A to G. Small triangles (▼▲), transmembrane region of VAMP and syntaxin. **c**, The four palmitoylated cysteines of SNAP-25. Axial (**c**) and vertical (**d**) projections of the motif arranged in an α -helical structure are shown: Glu or Asp residues are concentrated on one-third of the helix, contiguous to a cluster of hydrophobic residues in one-third of the helix.

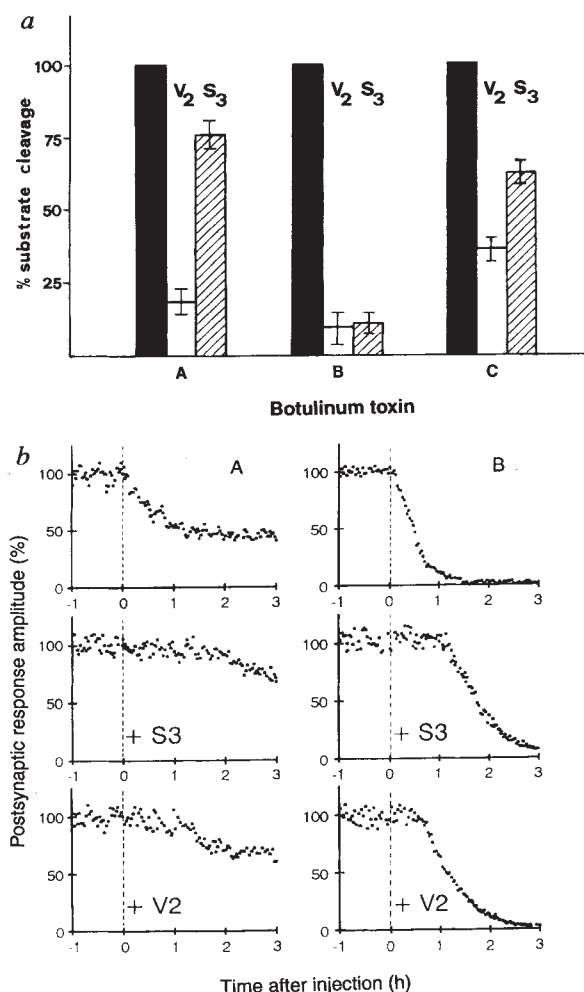


FIG. 2 The proteolytic activity (a) and blockade of neuroexocytosis (b) of botulinum neurotoxin A, B and C are inhibited by VAMP- and SNAP-25-derived peptides. a, Proteolytic activities of botulinum A (left), B (centre) and C (right), measured as before⁷, in the absence of added peptides, were taken as 100% (full bars) and compared with those found in the presence of 500 μ M V2 or S3 (open and hatched bars, respectively). b, Acetylcholine release of neurons of the buccal ganglion of *A. californica* was assayed as before⁷. Neurons were injected with botulinum A (left panels) or B (right panels), either alone (upper panels) or in neurons preinjected with S3 (middle row) or V2 (bottom panels). V2 and S3 alone did not affect acetylcholine release.

californica. V2 and S3 do not affect neuroexocytosis, but inhibit the blockade of acetylcholine release induced by botulinum A and B; C has little effect. Scrambled peptides have no effect on neurotransmitter release and on neuro-

toxin action. The neurotoxin recognition motif has to be exposed to the surface and, if it acts as an epitope, antibodies against the SNARE motif should recognize all three targets. Immunoblotting with affinity-purified polyclonal antibodies anti-V2 and anti-S3 recognize VAMP, SNAP-25 and syntaxin at the same time. Patterns of cross-reactivity are also found with some polyclonal antibodies raised against recombinant SNAREs. Parallel to this antibody cross-reactivity, one would expect a cross-recognition of the targets by the various neurotoxins. Indeed, botulinum A does not cleave VAMP, but inhibits VAMP proteolysis of botulinum B and tetanus neurotoxins. Conversely, tetanus neurotoxins and botulinum B neurotoxins inhibit the botulinum A cleavage of SNAP-25.

On the basis of these findings, we suggest that tetanus and botulinum toxins recognize their substrates via a double interaction with region A, a structural motif common to the three targets, and region B, the segment containing the peptide bond to be cleaved. B differs among the different neurotoxins and is located at variable distance from A in the primary structure of the three targets (Fig. 1). Available data^{1,7-10} indicate that region A is always amino-terminal to B. The present identification of a recognition site A distinct from the cleavage site B explains recent analysis of the minimal substrate segment still cleaved by the toxins⁸⁻¹⁰, although the relative contribution of A and B to specificity of binding and rate of proteolysis remains to be studied. The negative charges of the motif appear essential.

This report is, to our knowledge, the first example of a double recognition of a substrate by a metalloproteinase and parallels that of the serine-protease thrombin versus the thrombin receptor¹¹. It suggests a new method for the design of peptide inhibitors of clostridial neurotoxins of potential therapeutic value.

VAMP, SNAP-25 and syntaxin are essential to the process of vesicle docking-priming-fusion with the target mem-

brane¹⁻⁴. They possess multiple copies of a motif, not shared by any known neuronal protein. Among eukaryotic proteins, the motif is found also in yeast BPP4 and amino-end-recognizing proteins, as deduced from an inspection of the Swiss Protein databank with the FASTA program. This result suggests that VAMP, SNAP-25 and syntaxin are the sole neuronal targets of the proteolytic activity of clostridial neurotoxins. Moreover, it emphasizes the functional importance of this motif as a site that cannot be mutated under the selective pressure of neurotoxin action without loss of function. The physiological function of this segment is unknown, but a mutational scanning of the motif via site-directed mutagenesis and cell transfection will allow evaluation of its role in the exo-endocytosis cycle.

**Omella Rossetto, Giampietro Schiavo
Cesare Montecucco***

Dipartimento di Scienze Biomediche and
Centro CNR Biomembrane,
Università di Padova,
75 Via Trieste, 75 - 35121 Padova, Italy

Bernard Poulain, Florence Deloye
Laboratoire de Neurobiologie Moléculaire
et Cellulaire du CNRS,
91118 Gif-sur-Yvette, France

Luisa Lozzi

Dipartimento di Biologia Molecolare,
Università di Siena,
Le Scotte, 53100 Siena, Italy

Clifford C. Shone

Protein Toxins Section, CAMR,
Porton Down, Salisbury SP4 0JG, UK

*Author for correspondence.

Visual reference

SIR — It has been drawn to our attention that a paper written by Redies and others¹ on neuronal responses to subjective contours in the visual cortex of cats was not cited in our paper published last year² in which we suggested a new view of the primary visual cortex of macaque monkeys as an image-processing area. We did refer Redies *et al.*¹ in an interview for *Science* published in 1992, but in ref. 2 we cited only those papers that we felt were directly relevant to the points addressed therein. Nevertheless, we refer readers interested in the neurophysiology of subjective contours to the work of Redies *et al.*

David Grosf

Department of Ophthalmology,
Washington University, School of Medicine,
St Louis, Missouri 63110, USA

Michael Hawken

Robert Shapley
Center for Neural Science,
New York University,
New York 10003-6621, USA

- Redies, C., Crook, J. M. & Creutzfeldt, O. D. *Expl Brain Res.* **61**, 469-481 (1986).
- Grosf, D. H., Shapley, R. M. & Hawken, M. J. *Nature* **365**, 550-552 (1993).
- Winckelgren, I. *Science* **256**, 1520-1521 (1992).

- Montecucco, C. & Schiavo, G. *Molec. Microbiol.* **13**, 1-8 (1994).
- Huttner, W. *Nature* **32**, 3-4 (1993).
- Ferro-Novick, S. & Jahn, R. *Nature* **370**, 191-193 (1994).
- Söllner, T. *et al.* *Nature* **362**, 318-324 (1993).
- Rothman, J. & Warren, G. B. *Curr. Biol.* **4**, 220-233 (1994).
- Bennett, M. K. & Scheller, R. H. A. *Rev. Biochem.* **63**, 63-100 (1994).
- Schiavo, G. *et al.* *Nature* **359**, 832-835 (1992).
- Shone, C. C. *et al.* *Eur. J. Biochem.* **217**, 965-971 (1993).
- Yamasaki, S. *et al.* *J. Biol. Chem.* **269**, 12764-12772 (1994).
- Binz, T. *et al.* *J. Biol. Chem.* **269**, 1617-1620 (1994).
- Vu, T. H. *et al.* *Nature* **353**, 674-677 (1991).

Willow, titwillow, titwillow!

Steve Blinkhorn

The Bell Curve: Intelligence and Class Structure in American Life. By Richard J. Herrnstein and Charles Murray. *The Free Press*: 1994. Pp. 845. \$30, £25.

Measuring the Mind: Educational Psychology in England, c. 1860–1990. By Adrian Wooldridge. *Cambridge University Press*: 1994. Pp. 448. £45, \$69.95.

Race, Evolution, and Behavior: A Life History Perspective. By J. Philippe Rushton. *Transaction*: 1994. Pp. 334. \$34.95.

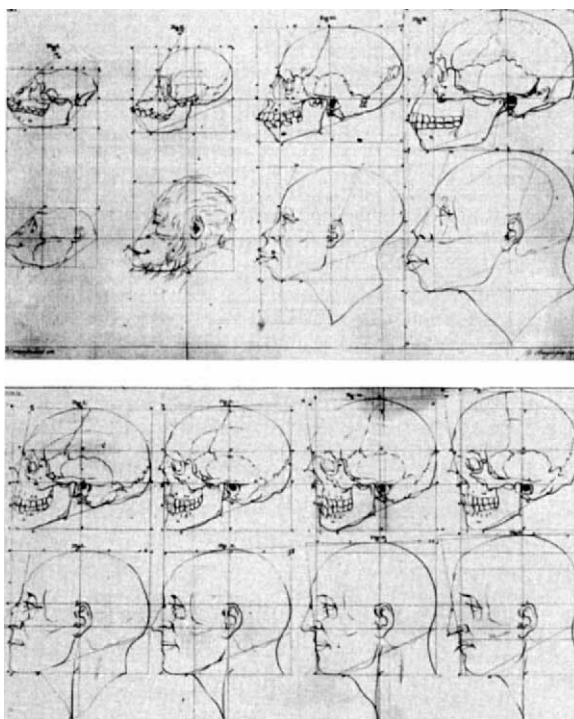
*'Is it weakness of intellect, birdie?' I cried,
'Or a rather tough worm in your little inside?'
With a shake of his poor little head he replied,
'Oh, willow, titwillow, titwillow!'*
W. S. Gilbert, *The Mikado* (1885)

WATCH out for the bandwagon, hold tight for a bumpy ride and take a long spoon with you. Intelligence testing is making a bid for intellectual respectability again, and the company is mixed. *The Bell Curve* has been heralded by an efficient public relations campaign that has cheerfully misrepresented its object, to the extent that I found it a much better book than I had been led to expect. Give a book a bad name and perhaps sales will soar. Its thesis is twofold: American society has developed in such a way that a self-perpetuating intellectual élite now dominates certain occupations, threatening social cohesion and throwing the lot of the less favoured into sharp relief; the driving force behind this stratification is none other than intelligence as conceived in traditional IQ terms. The first part of the thesis is, in fact, independent of the second, but each reinforces the flaws of the other.

It is worth saying at the outset that this is no red-necked rant, although no doubt the red-necked ranters of the world will think it a godsend. *The Bell Curve* assembles evidence that educational institutions and certain occupations have in the space of living memory become more selective and more specialized in the range of people recruited, and that IQ, rather than socioeconomic status for instance, is the variable that captures this shift. The effect of this change has been to introduce an unprecedented degree of segregation — physical, social and intellectual — into American society. The evidence is presented in a rather cool and unspectacular form, neatly documented with multivariate statistical analyses of a conventional kind reduced to a clear and comprehensible graphical format.

So the Something Must Be Done school of social pessimism has acquired a new sourcebook containing a litany of woes to be addressed by the very cognitive

élite that is its clearest symptom. Leonard Darwin's *The Need for Eugenic Reform* comes to mind as belonging to the same intellectual blood line, along with the rather less restrained writings of Raymond Cattell in the 1930s, although it has to be said that Herrnstein and Murray



Pieter Camper's 1792 study of racial types. Taken from *Nature's Body: Sexual Politics and the Making of Modern Science* by Londa Schiebinger (Pandora, £15.99).

present a better argued and better documented case.

The case is, nonetheless, fundamentally and irredeemably flawed. The more convinced you are of the hereditarian position with regard to IQ, of the genetic basis of individual psychological differences and of the virtues of intelligence tests as revealing underlying native differences in cognitive capacity, the less credence you will give to the notion that a self-perpetuating cognitive élite has emerged in society. And the more convinced you are that a self-perpetuating cognitive élite has emerged, the less credence you will give to the hereditarian position with regard to IQ.

For if heredity plays a large part in the determination of intelligence, then one expects it to operate according to the

same rules that it manifests elsewhere, working its magic by way of the genotype, not the phenotype. Unless we are all to turn suddenly Lamarckian, there is no reason to suppose that each generation will not continue to produce a good proportion of its most able individuals from less able parents, or that able, high-achieving parents will not continue to find some of their offspring a disappointment in the intellectual stakes.

Herrnstein and Murray do devote a whole page to this argument, which they tackle under the rubric of regression to the mean. Regression to the mean, they explain, is a statistical phenomenon, not a biological phenomenon. The IQs of children regress, on average, to the mean of the group to which they belong. Groups to which such children may belong include Blacks, Whites and Latinos. Where, we may ask, do such groups come from? Are they merely statistical groups exhibiting statistical phenomena? Then let us define them out of existence, for to do otherwise is to assume their importance. Are they biological groups exhibiting biological phenomena? Then the regression is real. Or are they groups defined on a complex biological and social history, distilled into categories for the convenience of social statisticians? Then we must take account of their origins in considering the measures we use.

In fact the groups Herrnstein and Murray should be interested in, given their primary thesis, are the cohort of the most able, considered with respect to their origins, and the children of the most able considered with respect to their destinations in society. I do not doubt for a moment that black Americans now have proportionately fewer high-achieving offspring than white Americans, but they have some and there is every reason to suppose that their achievements are as real as the whites'. There is a history of a narrowing of the gap between mean black and white IQs in North America, not perhaps a smooth history, but I know of no reason in principle why it should not continue.

So far I have played Herrnstein and Murray's game of alluding in an undifferentiated sort of way to IQ. This goes very much against the grain. So long as they are concerned to identify in broad terms what drives the social segregation they believe they have identified, it is tolerable enough. But when they switch horses and start interpreting more detailed research on intelligence into social mechanisms and finding reasons why research conducted with other ends in view should be taken as having implications for the order-

ing of society, it is time to call a halt. A quick canter around the varieties of theory underpinning the notion of intelligence is to be found in the introduction, but it does not inform the book proper. For instance, scholastic aptitude test (SAT) scores are not IQs, although they correlate well with IQs in homogeneous populations. The distinction between mental tests and scholastic tests was made clearly and firmly in the first quarter of this century by Cyril Burt, and was the focus of his campaigns for reform in the



"Europe Supported by Africa and America" by Blake, an allegory of colonial relations from Stedman's *Narrative of a Five Years' Expedition against the Revolted Negroes of Surinam* (1796). Taken from *Nature's Body* (see p. 417). The US edition was reviewed in *Nature* 366, 387 (1993).

English educational system.

In fact Herrnstein and Murray kidnap Spearman's notion of *g* and merrily ignore the care with which he sought to define its nature. They take *g* to be the general factor, or the first unrotated principal component, of any collection of tests that happen to be used in the studies they draw on. At other times, particular tests are taken as markers for *g* regardless of whether they have been shown to be so in the groups under investigation. But however well they may correlate, and often they correlate very highly, they are not identical and to treat them as identical begs the question.

To the casual reader, if any book with more than 800 pages ever gets casual readers, or to those who form their opinions from secondhand reports, the most striking conclusion is the suggestion that remedial programmes of one kind or another to aid the less fortunate groups in *The Bell Curve's* scheme of things are a waste of resources. People come to have more or less in the way of cognitive ca-

capacity through no fault of their own, and there is little point in trying to raise the IQs of weaker brethren. As a professional psychometrician, I find this suggestion extraordinarily myopic: biology not just as destiny but as doom. Intelligence exists only in its expressions, and these expressions are conditioned by many things.

We, and by that I mean you as part of any proposed cognitive élite, live in a built intellectual environment as surely as in a built physical environment. The symbols, metaphors and systems of relations that constitute that environment condition the way we think about intelligence. When 'environment' is contrasted with 'heredity' it is too easy to suppose that all it means is modern home comforts, decent nutrition and a reasonable number of years of schooling. But environment also has to cover induction into the world of the dominant culture and access to its ways. There is no way of building an intelligence test without making assumptions as to what is common culture in the target populations, and all tests access intelligence by way of acquired skills. The only reason people are inclined to attribute the position of American blacks to race rather than to generations of relative deprivation is that racial identity is worn on the face, not under layers of clothing.

To quote an earlier author: "To hereditary differences of race, sex and social class I have no space to allude. The main conclusion that can be drawn from experimental work is, I think, the following: innate group-differences exist; but they are small. Training and tradition account for the more conspicuous. The inborn mental differences between class and class, between nation and nation, and between women and men, taken on the average and in the gross, are swamped by the far wider differences among the individual members that make up any single group." So wrote Burt in 1923 in his presidential address to the psychology section of the British Association for the Advancement of Science.

Burt and his influence on English educational policy hold centre stage in *Measuring the Mind*, a gem of a book that ought to be required reading for all those who pontificate on his work, on the rise of the IQ test in England or on the ossified class system they are both supposed to have reinforced. It tells the story of the social experiments that radical and progressive thinkers such as Burt inspired, and of which I for one was a beneficiary. The author writes as an historian and his work has the mark of real intellectual distinction, if not of zealous proof-reading.

This is the first book by an outsider that I have read that captures the flavour and the dynamics of the arguments between traditionalists and radicals in the shaping of educational theory and practice, the influence of psychological thinking in turning the focus onto the child and its needs, and the political decisions that shaped the education of generations.

Wooldridge has succeeded in entering the minds of his dramatis personae and exploring their activities in the proper historical context. That this should seem surprising is a reflection mostly on the abysmal grasp of what was going on on the part of other authors with an axe to grind. Failure to appreciate the context, origins and purposes of the mental-measurement movement between, say, 1900 and 1940 in Great Britain is at the root of a great deal of nonsense, including some of the nonsense in *The Bell Curve*.

Those who view IQ testing, selection in education and differentiation of educational provision as essentially right-wing pro-Establishment measures are in for a rude shock. The eleven-plus examination taken by children in England and Wales at the age of ten or eleven for selection of suitable candidates for grammar schools was one of the most important engines of social mobility — introduced as it was in the aftermath of a war which itself helped dismantle class barriers — and opened a door of opportunity for many. At the very least it demonstrated that reserves of real ability were to be found among the poor, the ill-educated and the deprived.

How well that lesson has been learned is open to doubt. Access to better-quality schools in the state sector is rationed by house prices and catchment areas, and access to higher education, although wider, comes with a price tag that squeezes the not-quite-poor. This is the real bell-curve nightmare: the stemming of recruitment of the able offspring of the less well-off into the upper reaches of education and business.

Of course, the eleven-plus selection scheme was fairly seriously flawed, not least in that some education authorities placed the cut-off for grammar-school entry to match the availability of places, with results that would be comical if they were not so deplorably short-sighted when provision for the sexes was separate and unequal. Any system of cut-offs applied in the upper mid-range of a distribution is likely to misallocate some significant proportion of individuals, given the imperfect precision of the tests. The dilemma remains: how to provide optimal opportunity throughout the ability range while keeping schools reasonably sized and sensibly located.

If *Measuring the Mind* has the mark of intellectual distinction, *Race, Evolution and Behavior* is an intellectual jumble sale posing as an *haute couture* collection. It is

a frank attempt to rehabilitate the concept of race as a primary descriptive category explaining and integrating differences in everything from IQ to the angular elevation of the erect penis. It proceeds by skimming the surface of countless studies of grab groups, and what is to be found on that surface is not always cream.

The broad drift of Rushton's argument is that the Caucasoid and Mongoloid races of man have evolved against harsher environmental pressure than the Negroid, leading to a typical ordering of racial characteristics related to the relative emphasis on *r* (gamete production, mating behaviour and high reproductive rates) or *K* (parental care, resource acquisition, kin provisioning and social complexity) reproductive strategies. Guess at which end the blacks come?

The supposed relationship between cranial capacity and IQ gets quite an airing. First Rushton sets out to show that Caucasoids and Mongoloids have bigger heads than Negroids. Next he claims that the evidence shows that Caucasoids and Mongoloids do better on standard psychometric tests than Negroids. Then one is invited to the conclusion that blacks are less intelligent because they have smaller heads. The fact that women have smaller heads than men is brushed aside; perhaps we should correct for body size; or perhaps the excess neurons men have are related to the spatial and mathematical thinking at which men excel.

The temptation is to undertake a demolition job on Rushton's approach to the analysis of statistical data: there is plenty of scope, and it would be great fun, but I am sure there are plenty of others who can do it just as well as I. Much more insidious is the uncritical, not to say unthinking, credence Rushton and others lend to inferences from standard test scores. According to some of the data he presents, around half the population of Africa are sufficiently mentally subnormal to need supervision or care. It is the same mistake that Herrnstein and Murray make in *The Bell Curve*, attributing measurement properties to tests beyond what has been or could be shown, and not recognizing the extent to which their content — including so-called culture-fair or culture-free tests — reflects our built intellectual environment.

Race, Evolution, and Behavior does a major disservice to the serious study of the biological basis of behaviour and left me shuddering. It is time sociobiologists took a long hard look at the standards they adopt in evaluating psychometric evidence that suits their enterprise. And I should know: I take a size 8 in hats. □

Steve Blinkhorn is at Psychometric Research and Development Ltd, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK.

Buried treasures

Keith Stewart Thomson

Systematics and the Fossil Record: Documenting Evolutionary Patterns. By Andrew W. Smith. Blackwell Scientific: 1994. Pp. 223. £19.95 (pbk).

It is now almost 30 years since Hennig's cladistic methods of systematic analysis burst on the scene, bringing both turmoil and order to a field, and replacing *gestalt* with logic. Cladistics is based on a deceptively simple proposition — that in looking for order in biological (or other) diversity one should group together only those species that share uniquely derived features. 'Similarity' (the old way of looking at things) must therefore be rooted in true phylogenetic connection, not convergence or the sharing of retained primitive characters. The rigour of analysis forces both the quantity and quality of data. It requires comparability of datasets.

As a result of this approach, order and pattern are now emerging from even the most intractable groups of organisms. (It is not by chance that Hennig was an entomologist.) But an odd result has developed. Simply speaking, there have been two main reasons for engaging in systematic studies of any group. The first is to find the pattern of relationships among

the species — "A is more closely related to B than to C". The second is to find evolutionary relationships, which fundamentally means the biblical "A begat B, who begat C". Both sorts of information have been thought, at various times, to be essential to analysis of a whole range of subjects — biogeographical relationships, for example. Cladistics, however, is capable of analysing only the first of these two. It determines 'sister-group' relationships but is logically incapable of determining 'ancestor-descendant' relationships. In the branching diagrams that form the stock-in-trade of cladistics, none of the nodes is named. So, ironically, what is called phylogenetic systematics produces not what we used to call 'phylogenies' but, rather, metaphylogenies ('patterns', for convenience).

Historically, evolutionary analysis and particularly palaeontology have been a mixture of analysis and narrative: logic and story-telling. Partly under the influence of cladistics and the quality of data it produces, palaeontology has moved away from the narrative mode. But two controversial problems remain: how can one in practice combine the sorts of data that palaeontology produces with those from recent organisms? And, in principle, is palaeontology even necessary to the now-misnamed phylogenetic systematics, or simply a distraction at best?

Smith's book is not only a superb



BABY bouncer — a juvenile Australian red kangaroo. On leaving the pouch, a young kangaroo will stay close to its mother, continuing to nurse until it is more than an year old. Taken from *Mitsuaki Iwago's Kangaroos*, a magnificent collection of colour pictures by an award-winning photographer whose work has appeared in *National Geographic* and *Life* magazines. Chronicle Books, \$35 (hbk), \$19.95 (pbk).

primer of cladistic methodology but also an argument for the essential role of palaeontology in pattern reconstruction, and vice versa. Smith forthrightly shows that the problems presented by fossil data represent true opportunities. Particularly perceptive are his discussions of 'species' in the fossil record, the meaning and use of higher taxa, and the role of biostratigraphical data in pattern analysis. His method is to combine beautifully worked examples (principally from the broad field

of invertebrate palaeontology) with straightforward exposition. Almost everything in this field is still controversial, but this is a book one can imagine going into several editions over the years and becoming the student's best friend. □

Keith Stewart Thomson is president of the Academy of Natural Sciences, 1900 Benjamin Franklin Parkway, Logan Square, Philadelphia, Pennsylvania 19103-1195, USA.

Shooting down the star of Bethlehem



WHAT was the Star of Bethlehem? Did it really exist? Or was it a device used by the gospel writers to add portent to the story of the birth of Christ? If it is more than simply a legend, what did the Wise Men really see in the sky?

These are the questions with which planetaria have been entertaining visitors for many years. A show that opened on 18 November at the new Munich planetarium has a different way of telling the tale, making the most of its high-tech computerized equipment, more advanced than that at any other planetarium.

Visitors will not get any firm answers from the show: biblical records are too inaccurate to give a precise date for the birth of Christ. But with other historical and ancient astrological documents that the planetarium's director, Thomas Krauppe, has brought into play, they are provided with a few clues.

There are some certainties, however: astronomers can describe the sky exactly as it was 2,000 years ago, and the planetarium's star-making equipment can display it. So visitors are taken through a cosmic detective story, which reveals how historians came to the conclusion that Christ was probably born between 6 bc and 8 bc. They learn about the science behind the heavenly bodies and are

encouraged to draw conclusions about the nature of the Star as evidence accumulates. And they are invited to indicate their conclusions by pressing buttons on the arms of their seats.

"Do you think the Star was a comet, a shooting star or a supernova?" visitors are asked. The majority support the notion of a comet, prompting the voice-over to comment: "An interesting result". "Do you think the Star was a comet or a convergence of planets?" it later asks, when the scientific evidence for each possibility has been described and displayed on the huge night sky projected above the audience. "An interesting result", repeats the impassive voice, when most people have switched allegiance to the notion of a conjunction of Jupiter and Saturn, which, they have now learnt, took place in 7 bc.

Aimed primarily at young teenagers, the show is didactic, absorbing and aesthetically splendid. Alexander van Bubenheim, a Munich-based musician who has worked with Bruce Springsteen and Neil Young, has written an impressive accompanying score.

"Der Stern von Bethlehem" is showing daily in the planetarium at the Deutsches Museum in Munich, Germany, until 15 January 1995.

Alison Abbott

For reference

Companion Encyclopedia of Psychology, Volumes 1 & 2 edited by Andrew M. Colman (Routledge, pp. 1,356, £150). Aims to provide "authoritative and in-depth reference material on all major branches of psychological research and professional practice". The book contains more than 60 articles by some 70 workers, and is organized into 13 sections covering: the definitions of psychology and its historical background; biological aspects of behaviour; sensation and perception; cognition; learning and skills; emotion and motivation; individual differences and personality; developmental psychology; social psychology; abnormal psychology; special topics, including parapsychology, gender issues and health; research methods and statistics; and the professions of psychology. Complete with glossary and multi-level index.

Encyclopedia of Virology, Volumes 1-3 edited by Robert G. Webster and Allan Granoff (Academic Press, pp. 1,600, £300). Claims to be the "largest single reference source of current virological knowledge... the first to bring together all aspects of the subject for a wide variety of readers". With 270 alphabetically arranged "mini-review" articles, the text covers biological, molecular and medical topics concerning viruses in animals, insects, plants and bacteria and is accompanied by more than 250 illustrations. A cross-referencing system links related articles, and comprehensive reading lists at the end of each article allow access to the primary literature.

The Hutchinson Dictionary of Scientific Biography edited by Roy Porter (Hutchinson, pp. 891, £50). Previously published in six volumes, this new updated edition now contains more than 1,200 profiles (averaging 850 words) of key figures, an expanded glossary of 1,800 scientific terms, lists of Nobel prizewinners and over 80 diagrams illustrating notable experiments and key discoveries. There are also seven chronological reviews of the main scientific disciplines, which place individual scientists' achievements (and misconceptions) in the context of their own time.

Larousse Dictionary of Scientists: From the Pioneers of Science to the Innovators of Modern Research edited by Hazel Muir (Larousse, pp. 595, £30). Some 30 British academics have contributed more than 2,200 entries (each of around 300 words). The entries, which are thoroughly cross-referenced, are also interlinked with indexes of discoveries, research topics and Nobel prizewinners.

Massive iceberg discharges as triggers for global climate change

Wallace S. Broecker

Observations of large and abrupt climate changes recorded in Greenland ice cores have spurred a search for clues to their cause. This search has revealed that at six times during the last glaciation, huge armadas of icebergs launched from Canada spread across the northern Atlantic Ocean, each triggering a climate response of global extent.

IN 1988 Hartmut Heinrich published an article summarizing his study of a curious set of sedimentary layers in cores from the Dreizack seamounts in the eastern North Atlantic¹. His conclusion was that these layers record the melting of six great armadas of icebergs which flooded the northern Atlantic during the last glaciation. Each of these bursts produced a prominent sediment layer rich in Canadian-derived ice-rafted debris and poor in foraminifera shells. When Heinrich's paper was published, the attention of the palaeoclimate community was directed towards the climate response to the Earth's orbital cycles, so his discovery went largely unnoticed. Only when a nearly identical record was published² from a core located several hundred kilometres away, with a claim that the fresh water generated by the melting of these ice armadas might have disrupted deep-water formation in the northern Atlantic, did the palaeoclimate community recognize the possibility that Heinrich's layers were more than a curiosity. Suddenly attention was refocused on the coupling between surging ice sheets and northward heat transport by the Atlantic's thermohaline circulation.

It was already known from the Greenland ice cores^{4,45} that during the last glacial period the northern Atlantic region experienced repeated large and abrupt climate changes. So far only one mechanism—changes in ocean circulation—has been put forward to explain these jumps. The idea behind this mechanism is that high-latitude climate is strongly influenced by heat released from the sea. The amounts and routings of this heat are closely tied to the global pattern of thermohaline circulation. One such route, a conveyor-like circulation operating in the Atlantic, is particularly important. The heat it carries maintains the anomalously warm climate enjoyed by western Europe. The abrupt climate changes revealed in the Greenland record seem to be telling us that the conveyor has turned on and off. Modelling efforts reveal that such changes can be triggered by inputs of fresh water. One potential source of fresh water is that released by the melting of icebergs, such as the discharges recorded in Heinrich's layers (now known as Heinrich events).

Here I review the evidence that Heinrich events are indeed connected with rapid climate variations in the North Atlantic region. The timing of the events is in striking coincidence with the pattern of climate fluctuations documented from ice cores. But the mechanism that drives the events remains a matter of debate. One possibility is that the iceberg discharges reflect an internal oscillatory feature of the dynamics of the Laurentide ice sheet, from which the icebergs come. Alternatively, Heinrich events might represent a response to, rather than a cause of, climate change—external factors may induce global cooling, causing the ice sheets to surge. Although the true picture is likely to emerge only when our understanding of ocean circulation, ice-sheet dynamics and detailed regional palaeoclimate has improved substantially, there can be no doubt that Heinrich events provide a vivid illustration of the way in which the cou-

pled influence of the atmosphere, geosphere and cryosphere may be a dominant control on the Earth's climate.

Anatomy of Heinrich layers

Although glacial sediments of the northern Atlantic are uniformly rich in ice-rafted debris, that contained in Heinrich layers differs in three distinct ways. First, 20% of the sand-sized fragments are detrital limestone (a constituent nearly absent in ambient glacial sediment)³. Second, the clay-sized minerals in Heinrich layers have twofold higher K–Ar ages (~1,000 million years) than those for their ambient glacial equivalents^{4,6}. Third, Heinrich layers are devoid of the basal-derived clay minerals so abundant in ambient glacial sediment^{4,6}. Coupled with the observation that the detrital layers thin by more than an order of magnitude from the Labrador Sea to the European end of the 46° N iceberg route³, these composition differences point strongly to a Canadian origin for the Heinrich ice armadas.

Surprisingly, the dominant feature of Heinrich layers is not a greatly enhanced abundance of ice-rafted debris, but rather a dearth of foraminiferal shells^{2,3}. The abundance of these shells drops from its usual glacial value of thousands per gram to hundreds or less per gram in the Heinrich layers. Dilution with ice-rafted debris cannot be the sole cause of this reduction because the geographical extent of sediment sparse in foraminifera exceeds that of the Canada-derived detritus^{2,3}. Rather, the low abundance of foraminifera must reflect, in part, a dramatic decrease in marine productivity. Heinrich events occurred when the northern Atlantic was at its coldest, as signalled by the geographical extent of the polar foraminifera, *Neogloboquadrina pachyderma* (left coiling) which extended all the way to 40° N. The depleted ¹⁸O content in the few foraminifera present in Heinrich layers suggests the presence of a low-salinity lid³. These observations suggest episodes of extensive sea-ice cover similar to that of today's Arctic.

Timing of Heinrich events

Perhaps the most startling aspect of the timing of these events is their occurrence at major climate boundaries⁷. As shown in Fig. 1, Heinrich event number 6 marks the transition from the relatively warm northern Atlantic of the last interglacial (that is, marine stage 5) to the cold conditions prevailing during the last glacial (marine stages 4, 3 and 2). Heinrich event 1 marks the onset of the termination which brought the last glacial to a close. There is unpublished evidence (J. McManus, personal communication) of a seventh Heinrich event right at termination of the penultimate glaciation (that is, marine stage 6). The coincidence between Heinrich events and the major changes in the temperature regimen of the northern Atlantic leads to the speculation that fresh water released as the result of the melting of Heinrich's icebergs disrupted deep-water formation, thereby permitting switches between glacial and interglacial modes of thermohaline circulation.

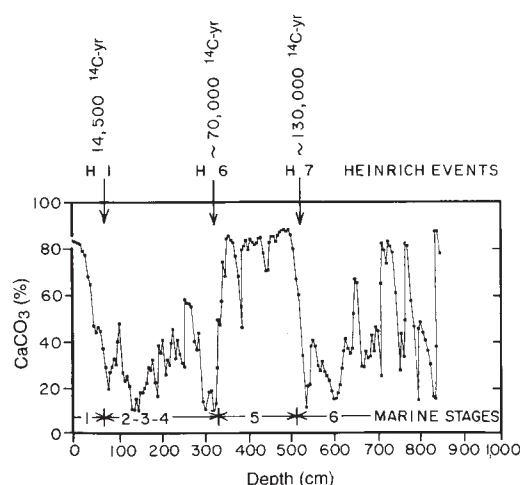


FIG. 1 Depth dependence of CaCO_3 content in core V28-82 from the northern Atlantic (50°N , 24°W) showing an alternation between coccolith-rich interglacial sediment (stages 1 and 5) and ice-rafted debris-rich glacial sediment (stages 2-4 and 6). Heinrich layers mark both the transitions from glacial to interglacial conditions and from interglacial to glacial conditions⁷.

Another intriguing feature of the timing of the Heinrich events is their spacing at intervals of roughly 10,000 years (ref. 8). This spacing suggests forcing associated with the alternation of the relative strength of the seasonality in the two polar hemispheres brought about by the Earth's precession. Perhaps this alternation influences the dynamics of the tropical atmosphere in such a way that the water-vapour burden of the temperate latitudes is changed. But the significance of this observation is not clear, given that the spacing between Heinrich events decreases through the course of the last glacial period from about 13,000 years to about 7,000 years.

Relationship to Dansgaard-Oeschger cycles

The glacial portion of the Greenland ice-core record is dominated by rectangular climate cycles averaging a few thousand years in duration⁹⁻¹¹. These so-called Dansgaard-Oeschger (D-O) cycles are characterized by abrupt jumps in temperature, dust content, ice accumulation rate, and concentration of methane and perhaps carbon dioxide. It is puzzling that Heinrich events are not prominent in this record. But as shown by Bond *et al.*¹², there is a tie between Heinrich events and D-O cycles. Bond *et al.* have shown that the ice armadas were launched during the most intense cold phase of a package of several D-O cycles, and that each Heinrich event was followed by a prominent warming which initiated a new package of D-O cycles (see Fig. 2). This bundling of progressing colder D-O cycles into subsets gives rise to what has come to be called Bond cycles.

The last of the D-O cold events is the Younger Dryas (YD). Its abrupt end ushered in the Holocene, a period of stable warm climate which has now lasted for about 11,500 calendar years (10,000 ^{14}C -years). The climate signal of the YD can be seen clearly in records from throughout the northern Atlantic basin. Bond's discovery³ of detrital carbonate in marine sediments of YD age from the northern Atlantic suggests that it may also have an affinity to the Heinrich events.

The global imprint

Evidence is rapidly accumulating that D-O cycles, Bond cycles and the YD all have global signatures (Fig. 3). Analyses of the air trapped in Greenland ice have revealed minima in methane concentrations associated with the cold phases of D-O cycles (including the YD event)¹³. As the major source for this gas during glacial time is thought to have been tropical wetlands, climate effects not directly attributable to changes in the north-

ward transport of heat by the Atlantic's conveyor must have been operative. Wet events in the pollen record from Lake Tulane in Florida appear to correlate with Heinrich events¹⁴. At least four maxima of the extent of Andean mountain glaciers in the lowlands of Chile have been identified by Denton and colleagues (personal communication). Radiocarbon ages for the last three of these maxima correspond within a few hundred years to the times of Heinrich events 3, 2 and 1. Denton and Hendy have also convincingly documented that mountain glaciers in the Alps of New Zealand reached an intermediate maximum precisely at the time of the onset of the YD¹⁵. Evidence of the YD has also been reported in foraminiferal records from the Sulu Sea¹⁶, pollen records from British Columbia¹⁷, lake levels in Africa¹⁸ and oxygen and hydrogen isotope ratios in Antarctic ice¹⁹. These data are summarized in ref. 20. So no longer can it be assumed that the influence of these events was restricted to the northern Atlantic basin, weakening the case that they are tied exclusively to the heat release to the atmosphere in association with the formation of deep water in the northern Atlantic²¹.

The trans-US wet event

A possibly important clue regarding the role of Heinrich armadas in perturbing ocean circulation comes from the sequence of events following the last of the six Heinrich outbursts. It seems that close to the time of this event, ice caps throughout the world reached prominent maxima²². Then shortly after Heinrich event 1, these glaciers went into rapid retreat (Fig. 4). This retreat is heralded by the basal sediments of the numerous small lakes in Switzerland's major Alpine valleys which document that by about 14,000 ^{14}C -years ago the valleys had become ice-free²³. However, pollen in the sediments of these lakes reveals that conditions remained sufficiently cold to suppress the re-appearance of trees²⁴. Then at about 12,700 ^{14}C -years ago, a sudden warming occurred, not only in Switzerland, but throughout the northern Atlantic basin. Climate proxies from oxygen isotope^{25,26}, beetle²⁷ and pollen²⁶ records show that within a period of a few decades, warm climates prevailed. This warming has been widely attributed to a turn-on of the Atlantic's thermohaline circulation²¹.

But this raises a difficult question. What transpired during the 1,400 or so years between the launch of Heinrich armada no. 1

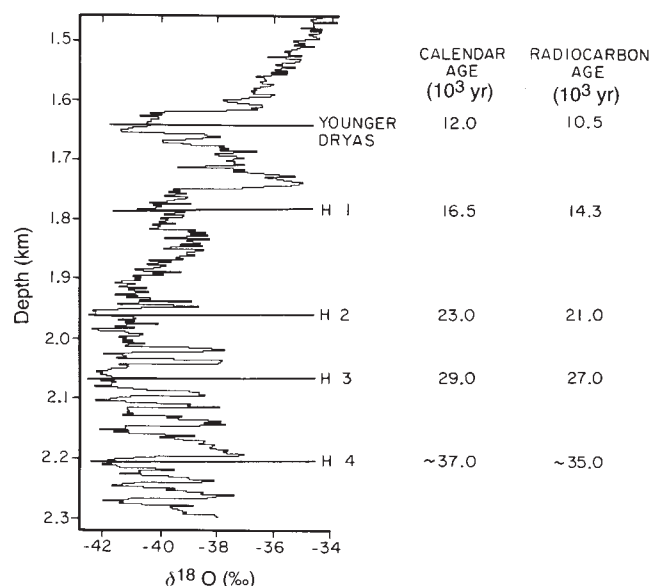


FIG. 2 Bond's placement of Heinrich events¹¹ in the Summit Greenland GRIP ice-core oxygen-isotope record⁹. Heinrich events occur in the last cold phase of a series of Dansgaard-Oeschger cycles and precede a major interstitial warm pulse¹¹. ($\delta^{18}\text{O}$ is the relative deviation of the $^{18}\text{O}/^{16}\text{O}$ ratio in a sample from that in standard mean ocean water.)

and the resumption of thermohaline circulation? A major clue comes from the global pattern of precipitation. Closed basin Lake Lahontan in the Great Basin of the western United States achieved its maximum size (Fig. 5) during this time interval and then underwent an abrupt major dessication just after 13,000 ^{14}C -years ago^{28,29}. As suggested by the pollen records from Lake Tulane, Florida¹⁴, and from Brown's Pond, Virginia³⁰, the southeastern United States was also unusually wet between 14,000 and 13,000 ^{14}C -years ago. This pattern of excess precipitation is similar to that experienced during El Niño periods³¹.

Ocean models offer a possible explanation for this curious interval. They show that sudden freshwater releases into polar regions can easily disrupt the existing pattern of thermohaline circulation. Whereas in some situations, the ocean immediately switches to an alternate thermohaline mode^{32,34}, in others, the model ocean lapses into what Manabe calls a "drop dead" mode with no deep-water formation^{35,36}. In either case, after a thousand or so years, the model's conveyor circulation can suddenly resume. Thus it is tempting to speculate that the interval between ~14,000 and ~12,600 ^{14}C -years ago was a "drop dead" ocean circulation phase which ended with a sudden rejuvenation of the Atlantic's conveyor. Of course, this line of speculation carries with it the implication that a lapse in deep-water formation somehow gives rise to wet events in North America akin to those characterizing El Niño intervals³¹. The nature of this connection remains a mystery.

Surges or fast-flowing ice streams?

The most obvious explanation for Heinrich events is that they result from surges of the ice stream that drained the Hudson

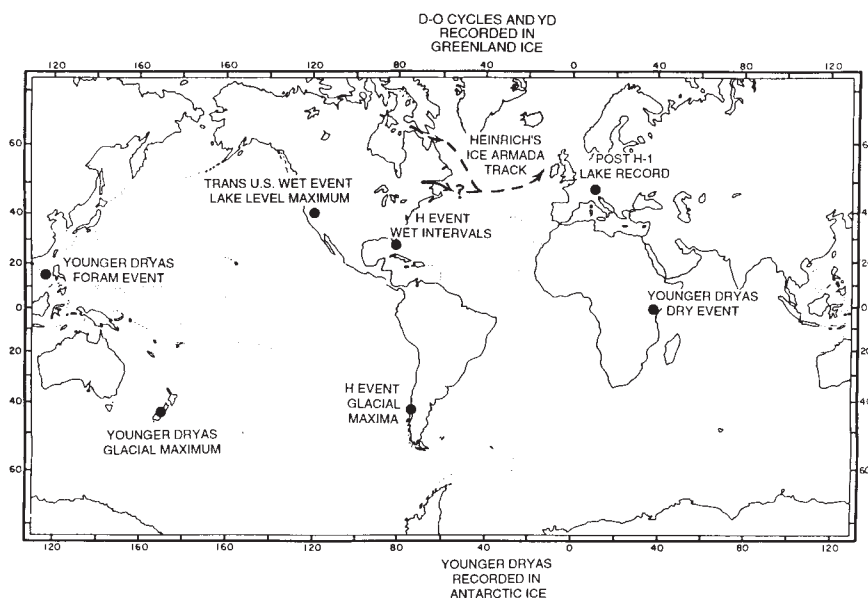


FIG. 3 Map showing the location of radiocarbon-dated records which, taken together, demonstrate the global signature of Heinrich events (H event) and the related Dansgaard-Oeschger (D-O), Bond cycles and the Younger Dryas.

Bay portion of the Laurentian ice sheet. Debris-laden basal ice issued forth from the Hudson Straits and gradually melted as it was carried by the prevailing currents across the Atlantic (along a track 10° wide in latitude centred at 46°N). Such an origin can adequately account for all we currently know about Heinrich layers. MacAyeal^{37,38} has proposed a binge-and-purge model for the Hudson Bay lobe of the Laurentian ice sheet to account for these surges. He and Alley have shown that this model has the added attraction of explaining how the large amount ($0.1\text{--}1.0\text{ km}^3$) of debris making up the Heinrich layers became frozen into the cap's basal ice^{39,40}.

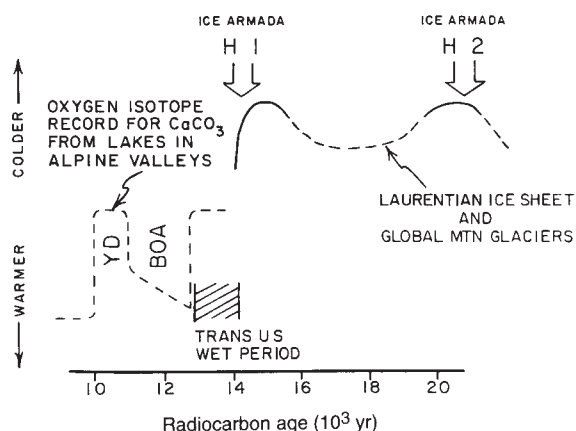


FIG. 4 Events of the last deglaciation: glaciers throughout the world achieved maxima close to 20,000 and to 14,500 radiocarbon-years (^{14}C -yr) ago. Based on the radiocarbon dates in hand, Heinrich armadas of icebergs were launched into the Atlantic close to the times of these glacial maxima. Following the second glacial maximum, the world's glaciers began a rapid retreat. In Europe this retreat opened the major Alpine valleys by 14,000 ^{14}C -yr ago allowing lakes to form. However, ~1,400 years passed before a second phase of warming occurred which allowed trees to colonize these valleys. This warming, which marked the onset of the northern Atlantic basin's Bolling-Allerod (BOA) warm period, occurred abruptly. The interval between the first Heinrich event (H 1) and the onset of the BOA corresponds to the time of an extreme wet interval across the southern tier of the United States (trans-US wet period).

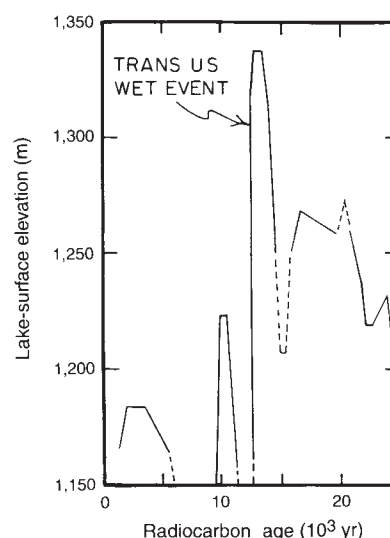


FIG. 5 The level of Lake Lahontan in the Great Basin of the western United States reached a brief late-Quaternary maximum, which dates between 12,700 and 14,500 ^{14}C -years ago (that is, 14,900–16,700 calendar years ago)²⁸. This age range is based on ^{14}C -ages for the organic material beneath desert varnish on shoreline cobbles²⁸, ^{14}C -ages for algal carbonate shoreline deposits^{28,29} and ^{230}Th – ^{234}U isochron ages on the same algal carbonates²⁹.

But the binge-and-purge hypothesis suffers from one major drawback—the timing of the surges would be controlled by internal properties of the ice sheet and not by external climatic forcing. If so, it is not so easy to understand what ties the advances of the Chilean mountain glaciers to the surges of the Laurentian ice sheet. As already stated, the two records are as near to synchronous as can be documented by existing radiocarbon measurements (within ± 300 years). Thus if the binge-and-purge theory is correct, each surge must somehow have sent out a signal to all the world's glaciers telling them to advance. The most likely carrier for this message would be the ocean's thermohaline circulation. Somehow its change would have to have led to a corresponding change in the Earth's temperature (perhaps via a change in the inventory of atmospheric water vapour).

Denton prefers to turn the situation around and invokes global temperature changes to drive both the glacial advances and generation of Heinrich icebergs. He calls on a global cooling to propel the rapid advances which pushed large quantities of Canadian ice into the sea, and allowed the icebergs to be carried far across the sea before melting. A problem with this interpretation is that because the response time of ice sheets to surface forcing is measured in tens of thousands of years^{41,42}, a close synchronism with rapidly responding mountain glaciers is not to be expected.

One important testable distinction exists between MacAyeal's and Denton's hypotheses. The former would restrict the release of Heinrich icebergs to the Hudson Straits whereas the latter would have them coming as well from the St Lawrence valley and even New England. There is a possibility that lead isotope measurements on individual feldspar grains, together with K–Ar measurements on individual amphibole grains would allow a determination of the source of Heinrich's icebergs. Their source would be defined unambiguously if they contained exclusively material from the older Precambrian Churchill Province (which underlies Hudson Bay) or if they also contained material from later Precambrian and Palaeozoic igneous bodies located to the south of the Canadian Archaean shield and also in eastern Greenland. Indeed, Gwiazda and Hemming have very recently obtained lead isotope measurements that clearly demonstrate that ice-rafted feldspars from Heinrich layer 2 are exclusively from older Precambrian rocks, whereas their equivalents in ambient glacial age sediment are derived from a wide range of terrains. This finding is consistent with MacAyeal's scenario.

Summary

So where does this leave us? First, the relationship between Bond cycles and D–O cycles remains a mystery. Bond cycles show up

in the glaciation of mountains in the Southern Hemisphere, pollen records from Florida, and iceberg records from the North Atlantic. Dansgaard–Oeschger cycles are found in global methane levels and in the Greenland ice-core records of temperature and dust. The YD displays characteristics of D–O and Bond cycles. As both phenomena seem to have their roots in the northern Atlantic, it is tempting to suggest the involvement of changes in the mode of thermohaline circulation, driven by the effect of fresh water on the formation of deep water in the northern Atlantic. Perhaps the D–O cycles are caused by a 'salt oscillator' operating in the Atlantic, with Heinrich icebergs disrupting the regular progression of these oscillations. Perhaps the changes in ocean circulation perturb the strength of upwelling in the equatorial zone, affecting the operation of the great tropical convective systems that load the atmosphere with much of its moisture. In this way, the climatic effects of ocean-circulation changes could be carried beyond the northern Atlantic into other parts of the Earth. If so, then one would expect that Denton's mountain glaciation record will eventually reveal lesser advances associated with individual D–O cold phases.

Although this all sounds feasible, I am not convinced. Lurking on the horizon is an as yet untapped source of information—the record of atmospheric $^{14}\text{C}/^{12}\text{C}$ ratios. As deep-water formation in the northern Atlantic is presently the dominant conduit for the transfer of ^{14}C from the upper ocean to the deep sea, any major changes in thermohaline circulation should measurably perturb the distribution of ^{14}C within the sea and hence $^{14}\text{C}/^{12}\text{C}$ ratio in atmospheric CO_2 . So far, we have suitable radiocarbon information (from independently dated tree rings, varved sediments and corals) only for the YD interval. These results clearly eliminate the possibility that thermohaline circulation was shut down during the Younger Dryas. Perhaps instead a shallower conveyor with a source region shifted to lower latitudes operated in its place⁴³. In any case the ^{14}C results loom as dark clouds on the horizon of an ocean-based model. Unfortunately, no atmosphere-based hypothesis has been proposed. Neither models nor theory provide a mechanism whereby the atmosphere on its own could jump from one quasi-stable mode of operation to another.

It must be pointed out that this subject is of more than academic interest. We are currently provoking the Earth's climate with a steady build-up of greenhouse gases. It would be prudent to find out what it might take to kick the system into one of its alternate modes of operation. Could this happen in the absence of greatly extended ice cover? □

Wallace S. Broecker is at Lamont-Doherty Earth Observatory, Columbia University, Route 9W, Palisades, New York 10964, USA.

- Heinrich, H. *Quat. Res.* **29**, 143–152 (1988).
- Broecker, W. et al. *Clim. Dyn.* **6**, 265–273 (1992).
- Bond, G. et al. *Nature* **360**, 245–249 (1992).
- Huon, S. et al. *Schweiz. miner. petrogr. Mitt.* **71**, 275–280 (1991).
- Huon, S. & Ruch, P. *Mar. Geol.* **107**, 275–282 (1992).
- Jantschik, R. & Huon, S. *Ecol. geol. Helv.* **85/1**, 195–212 (1992).
- Broecker, W. S., Bond, G. & McManus, J. in *Ice in the Climate System* (ed. Peltier, W. R.) 161–166 (NATO ASI Ser. 12, Springer, Berlin, 1993).
- Hagelberg, T. K., Bond, G. & deMenocal, P. *Paleoceanography* **9**, 545–558 (1994).
- Dansgaard, W. et al. *Science* **218**, 1273–1277 (1982).
- Johnsen, S. J. et al. *Nature* **359**, 311–313 (1992).
- Dansgaard, W. et al. *Nature* **364**, 218–220 (1993).
- Bond, G. et al. *Nature* **365**, 143–147 (1993).
- Chappellaz, J. et al. *Nature* **366**, 443–445 (1993).
- Grimm, E. C. et al. *Science* **261**, 198–200 (1993).
- Denton, G. & Hendy, C. N. *Science* **264**, 1434–1437 (1994).
- Linsley, B. K. & Thunell, R. C. *Paleoceanography* **5**, 1025–1039 (1990).
- Mathewes, R. W., Heusser, L. E. & Patterson, R. T. *Geology* **21**, 101–104 (1993).
- Roberts, N. et al. *Nature* **366**, 146–148 (1993).
- Jouzel, J. et al. *The Last Deglaciation: Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker, W. S.) 229–266 (Springer, New York, 1992).
- Peteet, D. et al. *EOS* **74**, 587–589 (1993).
- Broecker, W. S. *Oceanography* **4**, 79–89 (1991).
- Broecker, W. S. & Denton, G. H. *Quat. Sci. Rev.* **9**, 305–341 (1990).
- Schlichter, C. *Bull. Ass. fr. Étude Quaternaire* **2/3**, 141–145 (1988).
- Lotter, A. F. & Zbinden, H. *Ecol. geol. Helv.* **82/1**, 191–202 (1989).
- Siegenthaler, U., Eicher, U. & Oeschger, H. *Ann. Glaciol.* **5**, 149–152 (1984).
- Lotter, A. F. *Quat. Res.* **35**, 321–330 (1991).
- Atkinson, T. C., Briffa, K. R. & Coope, G. R. *Nature* **325**, 587–592 (1987).
- Benson, L. V. *J. Paleolimnol.* **5**, 115–126 (1991).
- Lin, J. C. et al. *Geochim. cosmochim. Acta* (in the press).
- Kneller, M. & Peteet, D. *Quat. Sci. Rev.* **12**, 613–628 (1993).
- Ropelewski, C. F. & Halpert, M. S. *Mon. Weath. Rev.* **115**, 1606–1626 (1987).
- Stommel, H. *Tellus* **13**, 224–230 (1961).
- Welander, P. in *Large-Scale Transport Processes in the Oceans and Atmosphere* (eds Anderson, D. L. T. & Willebrand, J.) 379 (NATO ASI ser. C, Dordrecht, Reidel, 1986).
- Marotzke, J. in *Oceanic Circulation Models: Combining Data and Dynamics* (eds Anderson, D. L. T. & Willebrand, J.) 501–511 (NATO ASI ser., Kluwer, Dordrecht, 1989).
- Manabe, S. & Stouffer, R. J. *J. clim.* **1**, 841–866 (1988).
- Manabe, S. & Stouffer, R. J. *Nature* **364**, 215–218 (1993).
- MacAyeal, D. R. *Paleoceanography* **8**, 767–773 (1993).
- MacAyeal, D. R. *Paleoceanography* **8**, 775–784 (1993).
- Alley, R. B. & MacAyeal, D. R. (abstr.) *EOS* **74**, 359 (1993).
- Alley, R. B. & MacAyeal, D. R. *Paleoceanography* **9**, 503–511 (1994).
- Whillans, I. M. *J. geophys. Res.* **86**, 4274–4282 (1981).
- Alley, R. B. & Whillans, I. M. *J. geophys. Res.* **89**, 6487–6493 (1984).
- Rahmstorf, S. *Nature* **372**, 82–85 (1994).
- Dansgaard, W., Johnsen, S. J., Clausen, H. B. & Langway, C. C. Jr in *Late Cenozoic Ice Ages* (ed. Turekian, K. K.) 37–56 (Yale Univ. Press, 1970).
- Oeschger, H. et al. in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 299–306 (Geophys. Monogr. 29, Am. Geophys. Un., Washington DC, 1984).

ACKNOWLEDGEMENTS. I am indebted to many people for helping to shape my thinking on this subject, including G. Bond, R. Alley, D. MacAyeal, S. Lehman, J. Andrews and H. Heinrich. I particularly thank G. Denton for regular correspondence about a whole range of aspects of this problem. Many of the ideas presented here have their origin in our dialogue, so in a sense, he is a shadow coauthor.

Positional cloning of the mouse *obese* gene and its human homologue

Yiying Zhang^{*†}, Ricardo Proenca^{*†}, Margherita Maffei[†], Marisa Barone^{*†}, Lori Leopold^{*†} & Jeffrey M. Friedman^{*†‡}

^{*} Howard Hughes Medical Institute, [†] The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The mechanisms that balance food intake and energy expenditure determine who will be obese and who will be lean. One of the molecules that regulates energy balance in the mouse is the *obese (ob)* gene. Mutation of *ob* results in profound obesity and type II diabetes as part of a syndrome that resembles morbid obesity in humans. The *ob* gene product may function as part of a signalling pathway from adipose tissue that acts to regulate the size of the body fat depot.

OBESITY is the commonest nutritional disorder in Western societies. More than three in ten adult Americans weigh at least 20% in excess of their ideal body weight¹. Increased body weight is an important public health problem because it is associated with type II diabetes, hypertension, hyperlipidaemia and certain cancers². Although obesity is often considered to be a psychological problem, there is evidence that body weight is physiologically regulated³.

The molecular pathogenesis of obesity is unknown. To identify components of the physiological system controlling body weight, we have applied positional cloning technologies in an attempt to isolate mouse obesity genes. Five single-gene mutations in mice that result in an obese phenotype have been described³. The first of the recessive obesity mutations, the *obese* mutation (*ob*), was identified in 1950⁴. *ob* is a single gene mutation that results in profound obesity and type II diabetes as part of a syndrome that resembles morbid obesity in humans⁵. Neither the primary defect nor the site of synthesis of the *ob* gene product is known. Cross-circulation experiments between mutant and wild-type mice suggest that *ob* mice are deficient for a blood-borne factor that regulates nutrient intake and metabolism⁶, but the nature of this putative factor has not been determined.

We report the cloning and sequencing of the mouse *ob* gene and its human homologue. *ob* encodes a 4.5-kilobase (kb) adipose tissue messenger RNA with a highly conserved 167-amino-acid open reading frame. The predicted amino-acid sequence is 84% identical between human and mouse and has features of a secreted protein. A nonsense mutation in codon 105 has been found in the original congenic C57BL/6J *ob/ob* mouse strain, which expresses a twentyfold increase in *ob* mRNA. A second *ob* mutant, the co-isogenic SM/Ckc-^{Dac}*ob*^{2J}/*ob*^{2J} strain, does not synthesize *ob* RNA. These data suggest that the *ob* gene product may function as part of a signalling pathway from adipose tissue that acts to regulate the size of the body fat depot.

For the positional cloning of mutant genes from mammals, it is necessary first to obtain genetic and physical maps, then to isolate the gene and detect the mutation. Here we describe the successful use of this approach to identify the *ob* gene.

Genetic and physical mapping of *ob*

The first *ob* mutation (carried on the congenic C57BL/6J *ob/ob* strain) was found proximal to the *Microphthalmia* (*Mt*) and *waved-1* (*wav-1*) loci on proximal mouse chromosome 6 (ref. 7); a second co-isogenic allele of *ob* has been identified in the SM/Ckc-Dac mouse strain (S. Lane, personal communication).

We previously positioned *ob* relative to a series of molecular markers on mouse chromosome 6 and mapped the *ob* gene close to a restriction-fragment length polymorphism (RFLP) marker, D6Rck 13, derived from chromosome microdissection^{5,8}; we also found that *Pax4* in the proximal region of mouse chromosome 6 is tightly linked to *ob* (ref. 9). Both loci were initially used to type a total of 835 informative meioses derived from both interspecific and intersubspecific mouse crosses that were segregating *ob*. *Pax4* was mapped proximal to *ob* and was recombinant in two animals (111 and 420 in Fig. 1); no recombination between D6Rck13 and *ob* was detected among the first 835 meioses scored⁸.

To isolate the *ob* gene we cloned the DNA in the region of *Pax4* and D6Rck13 (Fig. 1), using both probes to start the construction of a physical map in the region of *ob*. Yeast artificial chromosomes (YACs) corresponding to *Pax4* and D6Rck13 were isolated and characterized. Centromeric and telomeric ends of each YAC were recovered, and ends mapping closer to *ob* used to screen for new YACs. YAC ends were recovered using either vectorette polymerase chain reaction (PCR) or the plasmid end-rescue technique^{10,11}. One of the ends (labelled (1) in Fig. 1) of a D6Rck13 YAC, 902A0653, was recombinant in animal 257, positioning *ob* between this YAC end and *Pax4*. We were unable to recover any YACs linking the ends of YACs yB1S4A5 and 902A0653 (labelled (2) and (3) in Fig. 1). Pulsed-field gel electrophoresis (PFGE) indicated that there was a ~70-kb gap separating the two YAC ends. To bridge this gap, we used both YAC ends to isolate a set of mouse P1 clones^{12,13}. Analysis of the ends of these P1 clones showed that they overlapped and that the gap in the YAC contig was ~70 kb. The size of the contig spanning the *ob* locus was 2.2 megabases (Mb) and *ob* was localized to the 900 kb between the distal end of YAC 903E1016 (labelled (5) in Fig. 1) and the distal end of YAC 902A0653 (labelled (1) in Fig. 1).

To position *ob* more precisely, we genotyped an additional 771 meioses derived from both a C57BL/6J *ob/ob* × DBA/2J intercross and backcross⁵. The typing of the intraspecific crosses required the development of informative single-strand length polymorphisms (SSLPs) for both D6Rck13 and *Pax4*. Sequencing of the *Pax4* gene itself revealed a microsatellite sequence, and an SSLP near D6Rck13, D6Rck39, was identified by sequencing cosmid subclones from YAC y902A0653 (a YAC isolated with D6Rck13). PCR amplification of genomic DNA with primers flanking these microsatellites revealed polymorphisms among the various progenitor strains of the genetic crosses. No additional recombinants between *ob* and *Pax4* were identified after genotyping the obese backcross and intercross progeny from the crosses to DBA mice. The genetic results indicated that D6Rck39 was distal to *ob* and recombined with *ob* in a single obese animal

‡ To whom correspondence should be addressed.

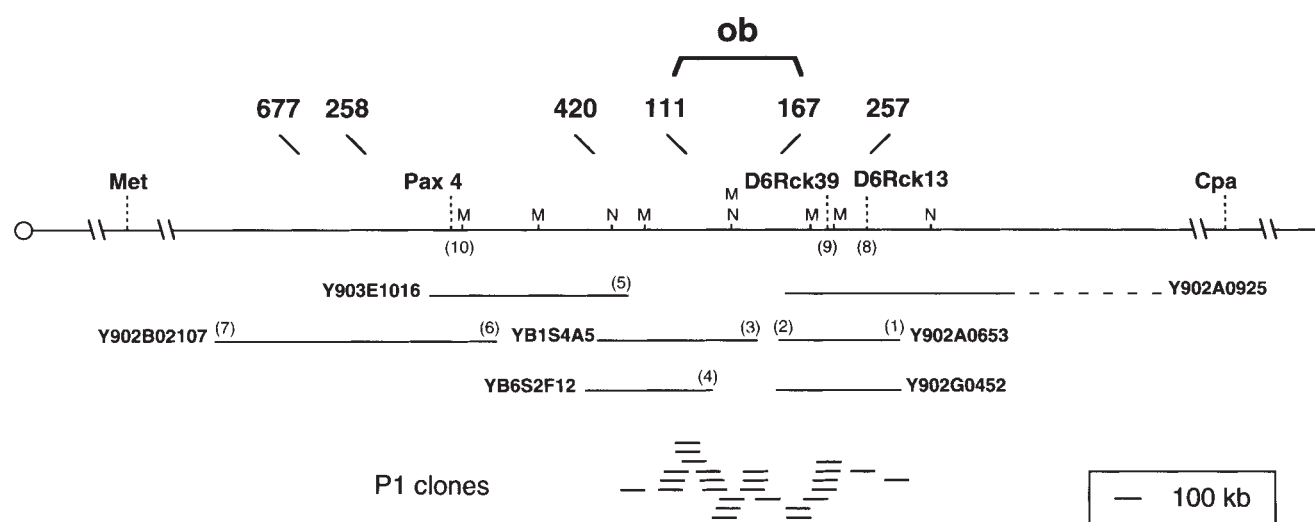


FIG. 1 Physical map of the mouse *ob* locus. A chromosome walk across the *ob* locus was completed in YACs and P1 clones using the flanking markers D6Rck13 and Pax4 as starting points. The position of the rare-cut restriction enzymes *Mlu*I (M) and *Not*I (N) are indicated. Numbers in bold type indicate designation of recombinant animals in the region of *ob* among the 1,606 meioses scored. Each of the ends of the YAC clones isolated with D6Rck13 and Pax4 were recovered using vectorette PCR and/or plasmid end rescue and used to type the recombinant animals. These ends were sequenced and used in turn to isolate new YACs. The resulting YAC contig is shown on the middle panel. One of the YACs in this contig, y902A0925, was chimaeric. Each of the probes used to genotype the recombinant animals is indicated in parentheses. On this basis, *ob* was localized to the interval between the recombination events in animals 111 and 167, (between end (5) and D6Rck39). Selected probes were also used to isolate bacteriophage P1 clones across the non-recombinant region. The resulting P1 contig is shown in the bottom panel. These P1 bacteriophage spanned a ~70-kb interval between probes (2) and (3) which could not be recovered in a YAC clone. The *ob* gene was identified in a P1 clone isolated using the distal end of YAC yB6S2F12(end 4).

derived from the C57BL6/J *ob/ob* × DBA/2J backcross, animal 167. This recombination occurred between D6Rck39 and the distal end of YAC yB6S2F12 (end (4) in Fig. 1) because a B × D polymorphism defined by this end indicated that it was non-recombinant in animal 167 as well as animals 111 and 420. Thus the distal end of YAC yB6S2F12 (end (4) in Fig. 1) failed to recombine with *ob* among the 1,606 meioses scored. As recombination in animal 111 was localized between this end and the distal end of YAC y902E1016 (end (5), which was non-recombinant in animal 420), the combined data from animals 111 and 167 placed *ob* in, at most, a 650-kb interval between D6Rck39 and YAC end (5). The exact size of the critical region could not be determined until the points of recombination in these animals were precisely localized. For reasons discussed later, fine mapping of these recombination events was not necessary. There was a total of 6 recombination events in this 2.2-Mb contig among 1,606 meioses. Therefore in the region of *ob*, 1 cM of genetic distance corresponded to ~5.8 Mb, a rate of recombination nearly threefold lower than average for the entire mouse genome¹⁴.

To facilitate the identification of genes in the region of *ob*, we isolated P1 clones using the ends of the YACs in this region as well as the ends derived from the initial set of P1 clones. Twenty-four P1 clones were used to construct a contig across most of the 650-kb critical region of *ob*.

Gene identification

Genes from this 650-kb interval were isolated using the method of exon trapping¹⁵. DNA from individual or pools of P1 clones was subcloned as *Bam*HI/*Bgl*II digests into the pSPL3 exon

METHODS. YAC clones were isolated by PCR screening or hybridization to high-density filters of available mouse YAC libraries^{38–40}. The YACs from ICRF begin with a 902 prefix. YAC clones were sized on a CHEF MAPPER (Bio-Rad). Restriction enzyme digestions were done according to the manufacturers recommendations. YAC ends were recovered using vectorette PCR and plasmid end rescue^{10,11}. P1 clones were isolated by sending specific pairs of PCR primers to Genome Systems Inc (St Louis, MO) who provided single picks of individual mouse P1 clones¹². P1 ends were recovered using vectorette PCR¹³. Cosmid subclones from YAC y902A0653 were isolated as described⁴¹. Primer selection and PCR amplification of simple sequence repeats were performed as described: initial denaturation at 94 °C for 3 min, 25 cycles of 94° for 1 min, 55° for 2 min and 72° for 3 min. Primer sequences for Pax4: Pax4F, 5'-GGAGGTAGAGATGGCAGCAG-3'; Pax4R, 5'-ACAGAAAGCAAGGAGGATTTC-3'; with a product size of 126 bp. Primer sequences for D6Rck39 were: 39GTF, 5'-GCACACTGACAGTGCCTTA-3'; 39GTR, 5'-TGTAACCTGGAATTGGGAGC-3' with a product size of 128 bp. The breeding and maintenance of the various mouse crosses have been described^{8,42}.

trapping vector. Briefly, each exon trapping product was sequenced and compared to all sequences in Genbank using the BLAST computer programme¹⁶. Putative exons were screened for the presence of corresponding RNA from a variety of tissues using northern blots and reverse-transcription PCR. Six genes were identified: four mapped within the 650-kb critical region of *ob* and two to outside this region.

One of the trapped exons, 2G7, was hybridized to a northern blot of mouse tissues and detected a ~4.5-kb RNA only in white adipose tissue (Fig. 2a). This exon was derived from a P1 isolated with the distal end of YB6S2F12 YAC (labelled (4) in Fig. 1). Even when autoradiographs were exposed for up to a week, no signal was detected in any other tissue, but we cannot exclude the possibility that this gene may be expressed elsewhere or below the level of detection in some tissues. Actin mRNA was detected in all samples (data not shown). Apparent adipose-tissue-specific expression was also seen after RT-PCR of RNA from a variety of tissues using specific primers from the 2G7 exon, with a strong signal being seen only in white fat and actin being detectable in all tissues tested (Fig. 2b). The high level of expression of 2G7 in adipose tissue suggested that this exon might be derived from the *ob* gene.

The level of expression of the 2G7 exon was then assayed in the two available *ob* strains by hybridization to northern blots as well as by RT-PCR. Northern blots showed that 2G7 RNA was absent in SM/Ckc- + ^{Dac}*ob*²³/*ob*²³ adipose tissue (Fig. 3a). This lack of 2G7 RNA in these animals was demonstrated by RT-PCR of the fat cell RNA, as shown by the absence of a signal after thirty cycles of amplification (Fig. 3b). As the *ob*²³ mutation is relatively recent and is maintained as a co-isogenic

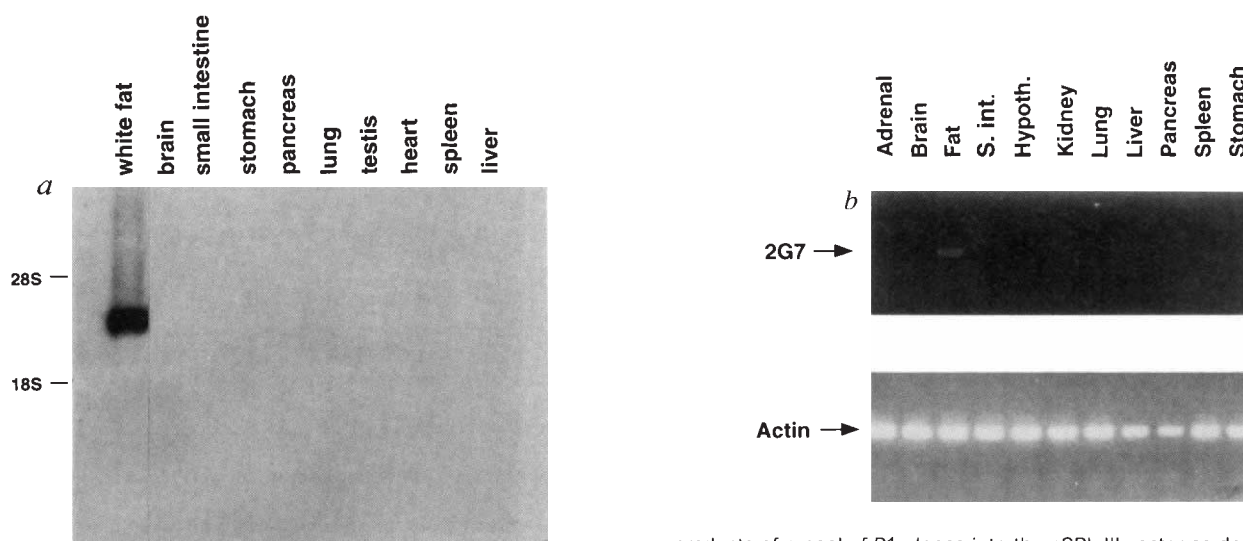


FIG. 2 Tissue distribution of the 2G7 transcript. *a*, Northern blot of total RNA (10 μ g) from various tissues probed with labelled 2G7 exon. The 2G7 exon was identified using exon trapping with DNA from a pool of P1 clones in the region of *ob*. This probe hybridized specifically to RNA from white adipose tissue. Autoradiograph signals appeared after 1-h exposure (24-h exposure shown here). The transcript migrated between 28S and 18S ribosomal RNA markers and is estimated to be $\sim$ 4.5 kb. *b*, Reverse transcription-PCR (RT-PCR) was performed with RNA from each of the tissue samples shown using primers specific for the 2G7 exon or actin. A positive signal was detectable only in white adipose tissue, even when PCR amplification was continued for 30 cycles. METHODS. *a*, Exon trapping was done by ligating *Bgl*II/*Bam*H1 digestion

products of a pool of P1 clones into the pSPL III vector as described¹⁵. Total RNA was prepared and electrophoresed in formaldehyde gels as in ref. 43. Northern blots were transferred to Immobilon and hybridized to radiolabelled probes as described⁴⁴. *b*, Reverse-transcription PCR reactions were performed using 100 ng total RNA⁴⁵. First-strand cDNA, prepared using random hexamers as primers, was PCR-amplified using primers derived from the 2G7 exon as well as mouse actin. The primers used to detect 2G7 were selected using the Primer program and were: 2G7F(5'CCAGGGCAGGAAAATGTG3'); 2G7R(5'CATCCTGGACTTTCTGGAT-AGG3'). The mouse actin primers were purchased from Clontech. PCR amplification was performed for 30 cycles with 94° denaturation for 1 min 55° hybridization for 1 min, and 72 extensions for 2 min with a 1-s autoextension per cycle. RT-PCR products were resolved in a 1.5% low-melting-point agarose (Gibco/BRL) gel run in 0.5 $\times$ TBE buffer.

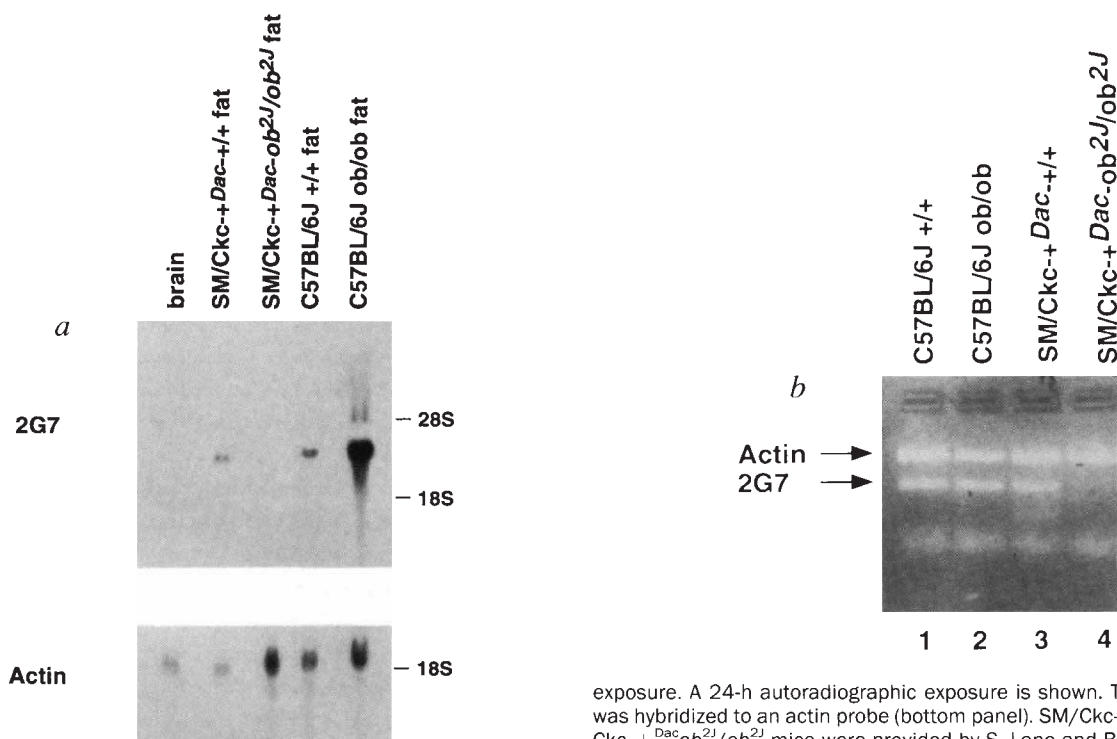


FIG. 3 2G7 expression in mutant mice. *a*, Northern blot of fat cell RNA isolated from obese and lean mice. The 2G7 exon was hybridized to northern blots with 10 μ g total RNA from white adipose tissue from each of the strains indicated. An approximately 20-fold increase in the level of 2G7 RNA was apparent in white fat RNA from the C57BL/6J *ob/ob* strain relative to lean littermates. There was no detectable signal in RNA from the SM/Ckc-+^{Dac}*ob*^{2J}/*ob*^{2J} mice even after a 2-week

exposure. A 24-h autoradiographic exposure is shown. The same filter was hybridized to an actin probe (bottom panel). SM/Ckc-+^{Dac} and SM/Ckc-+^{Dac}*ob*^{2J}/*ob*^{2J} mice were provided by S. Lane and B. Paigen of the Jackson Laboratory. C57BL/6J *ob/ob* mice were purchased from the Jackson Laboratory. *b*, RT-PCR of mutant RNA. Reverse-transcription PCR was performed for 30 cycles of amplification using 100 ng total fat RNA from each of the samples shown. In this experiment both the actin and 2G7 primer pairs were included in the same PCR reaction. The complete absence of 2G7 RNA was demonstrated in the SM/Ckc-+^{Dac}*ob*^{2J}/*ob*^{2J} adipose tissue. For RT-PCR reactions, 100 ng total RNA was used⁴⁵.

strain, these data indicate that 2G7 encodes an exon from the *ob* gene. This was confirmed by characterization of the mutation in C57L/6J *ob/ob* mice. Northern blots of adipose tissue RNA from C57BL/6J *ob/ob* mice showed a ~20-fold increase in the level of *ob* RNA (Fig. 3a), suggesting that the original *ob* allele (C57BL/6J *ob/ob*) was associated with a non-functional gene product and that the mRNA was increased as part of a possible feedback loop. This turned out to be the case (Fig. 4b) as the C57BL/6J *ob/ob* phenotype is the result of a nonsense mutation.

Sequence of the *ob* gene

The 2G7 exon was used to isolate a total of 22 complementary DNA clones from a mouse adipose tissue cDNA library. None of the cDNA clones extended more than 97 base pairs upstream of 2G7, whereas each extended a variable distance downstream. Sequencing of the cDNA clones revealed a methionine initiation codon in the 2G7 exon, with a 167-amino-acid open reading frame followed by a long 3'-untranslated sequence (Fig. 4). A potential Kozak translation initiation consensus sequence was present with an adenosine residue three bases upstream of the ATG¹⁷. One of the cDNA clones extended to the 5' end of the mRNA because its sequence was identical to that of the 5' RACE (rapid amplification of cloned ends) products of adipose tissue RNA (data not shown)¹⁸. A total of 2.9 kb of cDNA sequence is shown, most of which is 3'-untranslated sequence. A search for internal homology within the cDNA revealed a 50-base-pair (bp) direct repeat in both the 5' and 3' untranslated sequence; this sequence was not found in Genbank and there were no additional segments of internal homology.

Two classes of cDNA were found which differed by inclusion or exclusion of a single glutamine codon (underlined in Fig. 4). This residue is found in a position immediately 3' to the splice acceptor of the 2G7 exon. As the CAG codon of glutamine includes a possible AG splice-acceptor sequence, there appears to be slippage at the splice-acceptor site, with an apparent 3-bp deletion in a subset of the cDNAs¹⁹. This glutamine residue is located in a highly conserved region of the molecule but its significance is unknown.

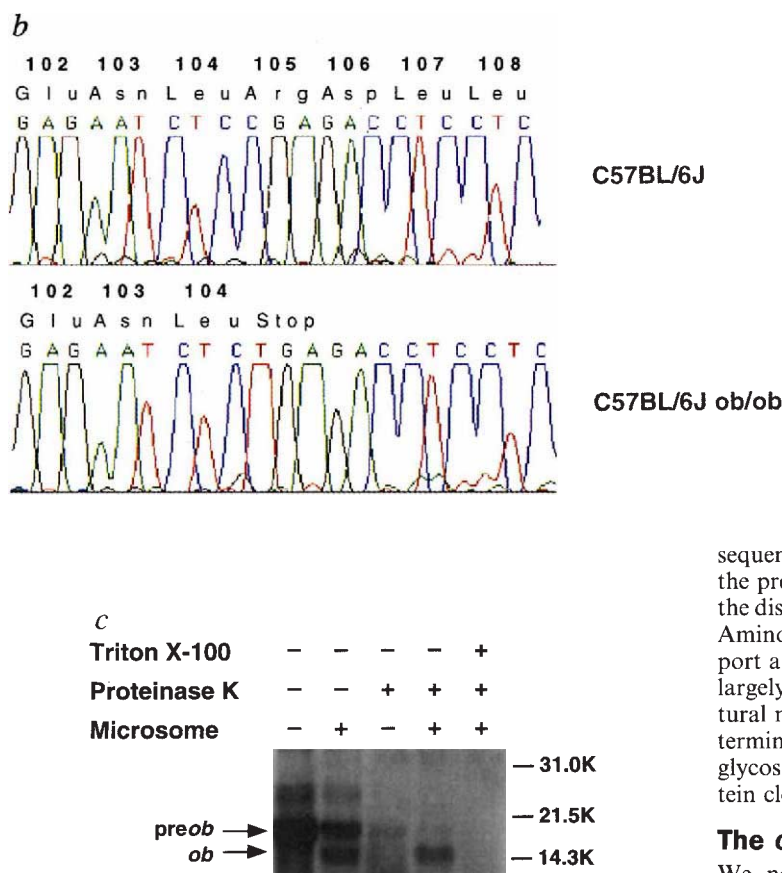
To identify the mutation in C57BL/6J *ob/ob* mice, RT-PCR products from the entire open reading frame (ORF) were prepared from adipose tissue RNA from this mutant and both strands sequenced. The coding sequences were identical apart from a C→T mutation in C57BL/6J *ob/ob* mice that results in a change of an arginine at position 105 to a stop codon (Fig. 4b). This amino acid is also underlined in Fig. 4a. This base change did not occur in genomic DNA from v/LE or C57BL/10 mice, the strains in which this mutation was carried before it was transferred to C57BL/6J (S. Lane, personal communication). DNA sequence changes from CGA to TGA are not uncommon as a result of the high mutation rate of methyl cytosine to thymidine²⁰. This *ob* mutation is presumed to inactivate the protein as the phenotype in both C57BL/6J and SM/Ckc-^{Dac}*ob²¹/ob²¹* mice is identical (S. Lane, personal communication; data not shown).

A database search of the *ob* protein using the BLAST program¹⁶ identified no significant homology to any sequences in Genbank. The predicted polypeptide was largely hydrophilic and had a putative N-terminal signal sequence²². Amino-acid sequence analysis using the *sigseq* computer program indicated a 98% likelihood that a signal sequence was present at the N terminus²¹ (underlined in Fig. 4a). The predicted signal sequence cleavage site is C terminal to an alanine at position 21. Computer analysis of the human protein (Fig. 6b) also suggests that it has an N-terminal signal sequence, so human and mouse *ob* may both encode secreted proteins. There is some divergence in the mouse and human signal sequences, but both give identical scores when analysed using *segseq*. To confirm the presence of a functional signal sequence, a human cDNA that included the entire ORF was subcloned into a pGEM vector (suitable mouse subclones were not recovered). Positive-strand human *ob* RNA

a

ATATTGCTGAGCTCAGGGAGTGAGGGCCCCACATTTGAGACAGTGAGCCCCAAGAAGAGG	60
GATCCCTGCTCCAGCAGCTGCAAGGTGCAAGAAGAAGAAGATCCAGGGGAGGAAATGTG	120
CTGGAGACCCCTGTGTGCGTTCTGTGGCTTTGGTCTATCTGTTATGTTCAAGCAGT	180
<u>W R P L C R F L W L W S Y L S Y V Q A V</u>	22
GCCTATCCAGAAAGTCCAGGATGACACCAAAACCTTCATCAAGACCATTTGTCACCGAGT	240
P I Q K V Q D D L T K T L I K T I V T R I	122
CAATGACATTTACACACGAGTCCGGTATCCGCAAGCAGAGGGTCACTGGCTTGGACTT	300
N D I S H T Q S V S A K Q R V T G L D F	62
CATTCTGGGCTTACCCCATCTGAGTTTGTCCAAGATGGACAGACTTGGCAGTCTA	360
I P G L H P I L S L S K M D I T L A V Y	82
TCAACAGGTCCTCACCAGCTGCCTTCCCAAAATGTGTCGACATAGCCAATGACCTGGA	420
Q Q V L T S L P S Q N V L Q I A N D L E	102
GAATCTCCGAGACCTCTCCATCTGCTGGCTTCCCAAGAGCTGCTCCCTGCCTCAGAC	480
N L R L L H L L S L S K M D I T L A V Y	122
CAGTGGCTCGCAAGCCAGAGAGCTGGATGGCGTCTGGAAGCCTCACTCTACTCCAC	540
S G L Q K P E S L D G V L E A S L Y S T	142
AGAGTGGTGGCTTTGAGCAGGCTGCAGGGCTCTCTGAGGACATTCTTCAACAGTTGGA	600
E V L A L S R L Q S L Q L D	162
TGTTAGCCCTGAATGCTGAAGTTTCAAGGCCACCGAGTCCCAAGAATCATGTAGAGGG	660
V S P E C *	167
AAGAAACCTTGGCTTCCAGGGGTCTTCAGGAGAAGAGGCCATGTGCACACATCCATCAT	720
TCATTTCTCTCCCTCCTGTAGACCACCCATCCAAAGGCATGACTCCACAATGCTTGACTC	780
AAGTTATCCACACAACCTTCATGAGCACAAGGAGGGGCCAGCTGCAGAGGGGACTCTCAC	840
CTAGTTCTTCAGCAAGTAGAGATAAGAGCCATCCATCCCTCCATGTCCACCTGCTCC	900
GGGTACATGTTCTCCGTGGGTACACGCTTCGCTGCGGCCACGAGAGGTTGAGGTAGGGA	960
TGGGTAGAGCCTTTGGGCTGTCTCAGAGTCTTTGGGAGCACCGTGAAGGCTGCATCCACA	1020
CACAGCTGGAACCTCCCAAGCAGCACAGATGGAAGCACTTATTTATTTATCTGCATTTC	1080
TATTTTGGATGGATCTGAAGCAAGGCATCAGCTTTTTCAGGCTTTGGGGGTGAGCCAGGA	1140
TGAGGAAGGCTCCTGGGTGCTGCTTTCAATCTATTGATGGGTCTGCCGAGGGCAAACC	1200
TAATTTTGGAGTGAAGGAGGTTGGGATCTTCCAAACAAGATCTATGCAGGTAG	1260
CGCTCAAGATTGACCTCTGGTGAAGTGGTGTCTTATTTGATGACTGACTTATCCAAAC	1320
ACGTTTGACAGGGCATTGCCGGGAGCATAGGCTAGGTTATTATCAAAGCAGATGAATTT	1380
TGTCAAGTGTAATATGTATCTATGTGCACCTGAGGGTAGAGGATGTGTTAGAGGGAGGGT	1440
GAAGGATCCGGAAGTGTCTCTGAATTACATATGTGTGGTAGGCTTTTCTGAAAGGGTGA	1500
GGCATTTTCTTACCTCTGTGGCCACATAGTGTGGCTTTGTGAAAAGGACAAAGGAGTTGA	1560
CTCTTTCCGGAACATTTGGAGGTGTACCAGGCACCTTTGGAGGGGTAAAGTACAGGCCT	1620
TTTGTGGCATATTTGCTGAGCTCAGGGAGTGAGGGCCCCACATTTGAGACAGTGAGCCCC	1680
AAGAAAGGGTCCCTGGTGTAGATCTCCAAGTTTGTCCAGGGTGTATCTCACAATGCGTT	1740
TCTTAAGCAGGTAGACGTTTGTATGCCAATATGTGGTTCTCATCTGATTGGTTTATCCAA	1800
AGTAGAACCTGTCTCCACCCATCTGTGGGGAGTTTGTTCAGTGGGAATGAGAAAT	1860
CAGTTAGCAGATGGTCTGAGCCCTGGGCCAGCAGCTGCTGAGGAAGTCCAGGGCCCCAG	1920
GCCAGGCTGCCAGAATTGCCCTTCGGGCTGGAGGATGAACAAGGGGCTTGGGTTTTTCC	1980
ATCACCCCTGCACCCATATGTACCATCAAATGGGGGGCAGATCAGTGAGAGGACACTTG	2040
ATGGAAGCAATACACTTTAAGACTGAGCAGAGTTTCTGTCTCAGCTCTGTCTGTGTGCTG	2100
TGAGCTAGAGAAGCTCACCACATACATATAAAATCAGAGGCTCATGTCCCTGTGGTTAG	2160
ACCTTACTCGCGCGGTGTACTCCACCACAGCAGCAGCCGACCGCTGGAAGTACAGTGCT	2220
GTCTTCAACAGGTGTGAAGAACCTGAGCTGAGGGTGACAGTCCCGAGGGGAACCTGCT	2280
TGCAGTCTATTGCATTTACATACCGCATTTTCAGGGCATTAGCATCCACTCCTATGGTA	2340
GCACACTGTGTACAATAGGACAAGGATAGGGGTTGACTATCCCTTATCCAAATGCTTG	2400
GGACTAGAAGAGTTTGGATTTTAGAGTCTTTTCAGGCATAGGTATATTTGAGTATATAT	2460
AAATAGAGATATCTTGGGATGGGGCCCAAGTATAAATCAAGTTCATTATATTTTCAT	2520
AATACCGTATAGACACTGCTTGAAGTGTAGTTTATACAGTGTTTTAAATAACGTTGTAT	2580
GCATGAAAGACGTTTTTACAGCATGAACCTGTCTACTCATGCCAGCACTCAAAACCTTG	2640
GGGTTTTGGAGCAGTTTGGATCTTGGGTTTTCTGTTAAGAGATGGTTAGCTTATACCTAA	2700
AACCAATATGGCAACAGGCTGCAGGACCAGACTGGATCCTCAGCCCTGAAGTGTGCCCT	2760
TCCAGCCAGGTCATACCTGTGGAGGTGAGCGGGATCAGGTTTTGTGGTGCTAAGAGAGG	2820
AGTTGGAGGTAGATTTTGGAGGATCTGAGGGC	2852

FIG. 4 a, Sequence of *ob* cDNA. 22 cDNA clones were isolated from a mouse white fat cDNA library and sequenced. A 97-bp 5' leader was followed by a predicted 167-amino-acid ORF and a ~3,700-bp 3'-untranslated sequence. A total of ~2,500 base pairs of the 3'-untranslated sequence is shown. Analysis of the predicted protein sequence using the *sigseq* computer



program suggests the presence of a signal sequence (underlined)²¹. Microheterogeneity of the cDNA was noted in that ~70% of the cDNAs had a glutamine codon at codon 49 and 30% did not. This amino acid is underlined as is the arginine codon that is mutated in C57BL/6J *ob/ob* mice (see *b*). After cleavage of the signal sequence, two cysteines remain at the C terminus of the molecule, possibly with formation of a disulphide bond. cDNA clones were isolated from a mouse adipose tissue library (Clontech)⁴⁶. Inserts were prepared directly from phage DNA or by PCR amplification of the insert using primers from the λ gt10 cloning site (Clontech). PCR products were prepared for sequencing after electrophoresis in low-melting-point agarose gels (Gibco/BRL) using agarase⁴⁷. DNA was sequenced manually or using an ABI 373A automated sequencer^{48,49}. 5' RACE was done after dG tailing of first-strand cDNA using terminal transferase followed by PCR amplification with the reverse complement of the 2G7F primer and oligo(dC)¹⁸. *b*, The C57BL/6J *ob/ob* mutation. RT-PCR was carried out using white fat RNA from C57BL/6J *+/+* and C57BL/6J *ob/ob* mice and primers from the 5' and 3' untranslated regions. PCR reaction products were gel-purified and sequenced using primers along both strands of the coding sequence. The C57BL/6J *ob/ob* mice had a C→T mutation which changed Arg 105 to a stop codon. This base change is shown as the output from the automated sequencer. The cDNA sequences of mutant and wild-type mice were otherwise identical. The wild-type cDNA sequence was also seen in genomic DNA from v/Le and C57BL/10 mice, the progenitor strains of the stock on which the C57BL/6J *ob/ob* mutation originally arose (data not shown). *c*, *In vitro* translation of *ob* RNA. A human *ob* cDNA was subcloned into the pGEM cloning vector (see Fig. 5*b*). Plus-strand RNA was synthesized using SP6 polymerase. The *in vitro* synthesized RNA was translated in the presence or absence of canine pancreatic microsomal membranes. An ~18K primary translation product was seen after *in vitro* translation. The addition of microsomal membranes (mb) to the reaction led to the appearance of a second translation product ~2K smaller than the primary translation product. The 16K product was resistant to proteinase K but was rendered protease-sensitive when the microsomal membranes were permeabilized by Triton X-100, indicating that a functional signal sequence was present. A human *ob* cDNA was cloned into the pGEM-3zf(+) plasmid. *In vitro* transcription was carried out using a Ribo MAX large-scale RNA production system with SP6 polymerase (Promega). The transcription mixture was used without further purification in a wheat-germ translation system (Promega) with or without 5 μ g of a canine pancreas microsomal membrane preparation⁵⁰ at 27 °C for 2 h. Proteinase K (100 μ g ml⁻¹) digests were performed on ice for 1 h with and without 0.1% Triton-X100. Translation products were analysed by SDS-PAGE.

was transcribed using sp6 polymerase and translated *in vitro* with and without canine pancreatic microsomal membranes⁵⁰. The primary translation product migrated with an apparent relative molecular mass of ~18K, consistent with that predicted from the cDNA sequence. Inclusion of microsomal membranes in the reaction inhibited translation ~5-fold, but about half of the *ob* primary translation product was truncated by ~2K in the presence of microsomal membranes, which suggests that the signal sequence is functional (Fig. 4*c*). To confirm that the *ob* protein had been translocated, we treated *in vitro* translation products with proteinase K, which caused complete proteolysis of the 18K primary translation product whereas the 16K processed form was unaffected, indicating that the translation product had been translocated into the microsomal lumen. Permeabilization of the membranes with Triton X-100 rendered the processed form protease-sensitive and is compatible with the hypothesis that *ob* is a secreted molecule. After signal-sequence cleavage, two cysteine residues would remain within the predicted protein, suggesting that the molecule may contain the disulphide bond characteristic of other secreted polypeptides. Amino-acid sequence and secondary structure prediction support a globular structure for the protein (data not shown). The largely hydrophilic amino-acid sequence had no notable structural motifs or membrane-spanning domains other than the N-terminal signal sequence. We find no consensus for N-linked glycosylation or dibasic amino-acid sequences indicative of protein cleavage in the predicted processed protein²².

The *ob*^{2J} mutation

We next explored the molecular basis for the mutation in SM/Ckc-+^{Dac} *ob*^{2J}/*ob*^{2J} mice. In these animals, the absence of 2G7 RNA was associated with an increase in the size of an ~9 kb *Bgl*/II fragment of genomic DNA in affected animals (Fig. 5*a*). Although the precise nature of this polymorphism is not established, the altered *Bgl*/II site maps ~7 kb upstream of the mRNA start site for the 4.5 kb *ob* RNA (data not shown), suggesting that this mutation may be the result of a structural alteration or sequence variation in the promoter. None of the other restriction fragments reach the promoter region. Nevertheless, this polymorphism is significant because it is always associated with the obese phenotype in a colony that segregates the *ob* 2J mutation (Fig. 5*b*). In this experiment, DNA from a total of six obese and five lean animals was Southern-blotted after *Bgl*/II digestion and probed with 2G7. In each case homozygosity for the larger of the polymorphic *Bgl*/II fragments was associated with an obese phenotype. Each of the lean animals was either homozygous for the smaller *Bgl*/II allele (*+/+*) or heterozygous (*ob/+*). As reported, no overt phenotypic differences were apparent between *+/+* and *ob/+* mice²³.

ob sequence conservation

The coding sequence of the *ob* gene was hybridized to genomic Southern blots of vertebrate DNAs (Fig. 6*a*). Even at moderate stringency, there were detectable signals in all vertebrate DNAs tested, demonstrating that *ob* is highly conserved among vertebrates. To determine the extent of this *ob* sequence conservation, we isolated and sequenced cDNA clones hybridizing to *ob* from a human adipose tissue cDNA library. The nucleotide sequences from human and mouse were highly homologous in the predicted coding sequence, but had only 30% homology in the available 5' and 3' untranslated regions. Homology of the human polypeptide apparent N-terminal signal sequence to the mouse equivalent region was slightly lower than in the rest of the molecule. The signal sequence is seen to be functional from the truncation of the primary translation product in the presence of canine pancreas microsomal membranes (Fig. 4*c*). Alignment of the predicted human and mouse amino-acid sequences (Fig. 6*b*)

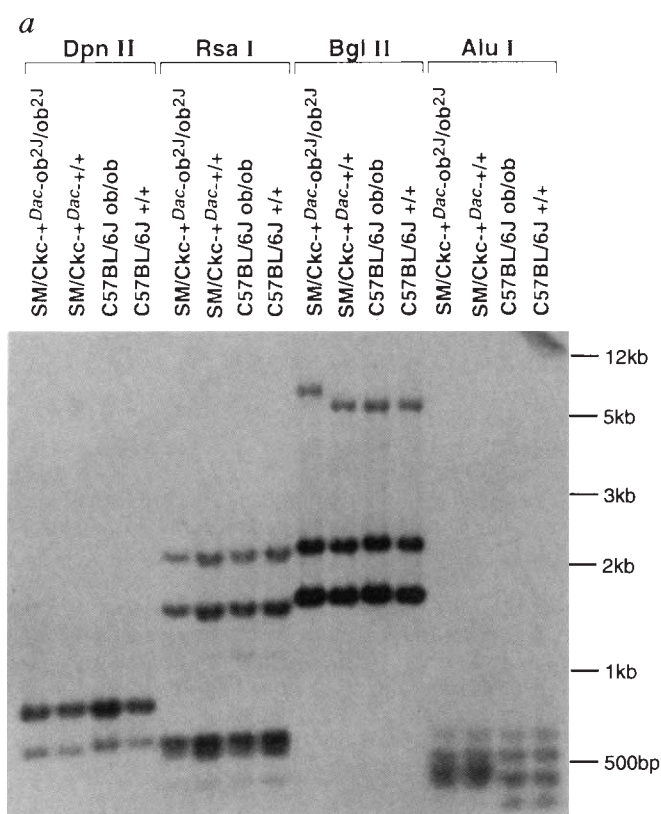
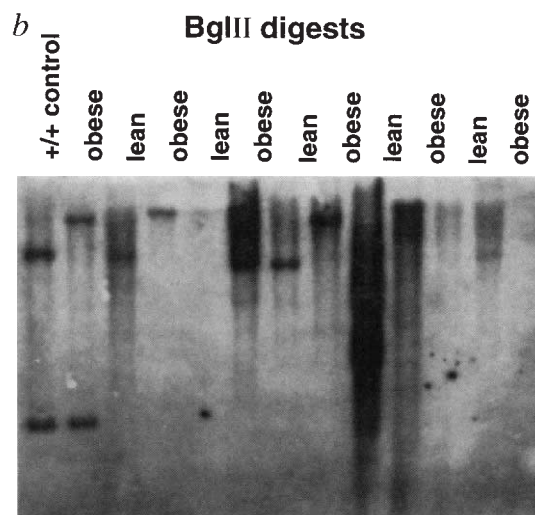


FIG. 5 The *ob^{2J}* mutation. a, Southern blots of genomic DNA from mutant animals. The 2G7 probe was hybridized to 5 μ g of restriction-enzyme-digested genomic DNA from each strain. Restriction digestion with *Bgl*II



revealed an increase in the size of a ~9 kb (the largest) *Bgl*II fragment in SM/Ckc-+*Dac**ob^{2J}/ob^{2J}* DNA. RFLPs were not detectable with any other restriction enzymes. Preliminary restriction mapping of genomic DNA indicated that the polymorphic *Bgl*II site is ~7 kb upstream of the transcription start site (data not shown). None of the other enzymes tested extend past the mRNA start site. Genomic DNA preparation and Southern blotting have been described⁸. b, Segregation of a *Bgl*II polymorphism in the SM/Ckc-+*Dac**ob^{2J}/ob^{2J}*. Six obese and five lean progeny from the same generation of the coisogenic SM/Ckc-+*Dac**ob^{2J}/ob^{2J}* colony were genotyped by scoring the *Bgl*II polymorphism in a. All of the phenotypically obese animals were homozygous for the larger allele of the polymorphic *Bgl*II fragment. The DNA in the control lane was prepared from an unrelated SM/Ckc-+*Dac**+/+* mouse, bred separately from the SM/Ckc-+*Dac**ob^{2J}/ob^{2J}* colony.

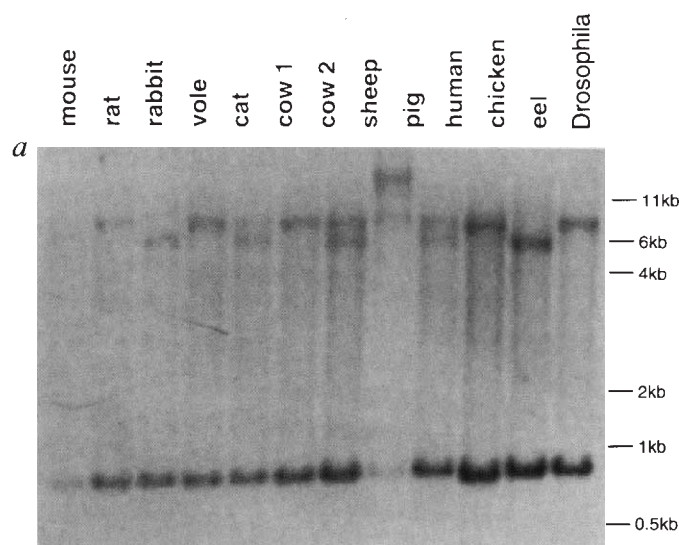


FIG. 6 Evolutionary conservation of *ob*. a, Cross-species hybridization of *ob*. Genomic DNA from each of the species shown was Southern blotted after *Eco*RI digestion and probed with *ob* cDNA. Hybridization signals were detectable in every vertebrate sample even after moderate-stringency hybridization. The cat DNA used in this experiment was slightly degraded. DNA was prepared from tissues or cell lines from the organisms shown. Southern blot hybridization was at 65 °C, with two washes in 2 $\times$ SSC/0.2% SDS at 65 °C for 20 min, and autoradiograph exposure was for 3 d using Kodak X-OMAT film. b, Sequence of human *ob* protein. The mouse *ob* gene was used as a probe to isolate human adipose tissue cDNA clones from a λ gt11 library. The sequence of human *ob* cDNA was highly homologous to that of the mouse cDNA in the predicted 167-amino-acid coding sequence. There was 30% homology

Mouse	MCWRFLCRFL WLWSYLSYVQ AVPIQKVQDD TKTLIKTIVT RINDISHTQS	50
Human	MHWGTLGGFL WLWPYLFYVQ AVPIQKVQDD TKTLIKTIVT RINDISHTQS	
Mouse	VSAKQRTVGL DFIPGLHPIL SLKMDQTLA VYQQVLTSLP SQNVLQIAND	100
Human	VSSKQKVTVGL DFIPGLHPIL TSKMDQTLA VYQQILTSLP SRNVIQISND	
Mouse	LENLRDLLHL LAFSKSCSLP QTSGLQKPES LDGVLEASLY STEVVALSRL	150
Human	LENLRDLLHV LAFSKSCHLP WASGLETLDS LGGVLEASGY STEVVALSRL	
Mouse	QGSLLQDILQQ LDVSPQC	167
Human	QGSLLQDMLWQ LDLSPGC	

in the 5' and 3' untranslated sequences in this tissue. Predicted amino-acid sequences of the human and mouse proteins have been aligned. Conservative changes are noted by a dash and non-conservative changes by an asterisk. The variable glutamine codon is underlined, as is the position of the nonsense mutation in C57BL/6J *ob/ob* mice. Overall there is 84% identity at the amino-acid level, although only seven substitutions were found between the valine at codon 22 (immediately downstream of the signal sequence overage) and the cysteine at position 117. cDNA clones were isolated from a human adipose tissue cDNA library (Clontech) as described for Fig. 4, except that PCR amplification of the phage insert was carried out using primers adjacent to the cloning site of λ gt11. PCR products were sequenced directly.

shows that there is 84% overall identity and more extensive identity in the N terminus of the mature protein, with only four conservative and three non-conservative substitutions among the residues between the signal sequence cleavage site and the conserved cysteine at position 117. As in mouse, 30% of the clones were missing a glutamine at codon 49.

Discussion

Obesity has been a focus of discussion²⁴, with a line of research dating back to Lavoisier and Laplace (1783) implying that energy balance—food intake versus energy output—is physiologically regulated^{25–27}. The site of this regulation has been the subject of nearly one hundred years of debate^{28–30}. Of the brain regions implicated in the regulation of feeding behaviour, the ventromedial nucleus of the hypothalamus (VMH) is considered to be the most important satiety centre in the central nervous system (CNS). The increase in body weight associated with VMH lesions is a result of both increased food intake and decreased energy expenditure²⁹.

The energy balance in mammals was therefore postulated to be controlled by a feedback loop in which the amount of stored energy is sensed by the hypothalamus, which adjusts food intake and energy expenditure to maintain a constant body weight^{31,32}. The nature of the inputs to the hypothalamus was unclear²⁷. According to the lipostasis theory³², the size of the body fat depot is regulated by the CNS, with a product of fat metabolism circulating in plasma and affecting energy balance by interacting with the hypothalamus. The glucostasis theory³³ suggested that the plasma glucose level is a key signal regulating energy stores. A third possibility is that body temperature is an important input to the CNS centres controlling food intake³¹. The inability to identify the putative signal from fat has hindered the validation of the lipostasis theory. Moreover, neither the glucostasis theory nor theories on the thermal regulation of food intake fully account for the precision with which energy balance is regulated *in vivo*. The possibility that at least one component of the signalling system circulates in the bloodstream was first suggested by Hervey³⁴, who showed that the transfer of blood from an animal with a VMH lesion across a vascular graft to an untreated animal (a parabiosis experiment) resulted in a reduction of food intake in the intact animal. The biochemical nature of the putative signal has remained elusive, although it has been suggested that *ob* is responsible for the generation of this blood-borne factor⁶.

Our results, particularly the evidence that the *ob* protein product is secreted, suggest that *ob* may encode this circulating factor. Evidence that the gene described here is *ob* includes the identification of a nonsense mutation in the C57BL/6J *ob/ob* strain and a genomic alteration in the SM/Ckc- + ^{Dac}*ob*^{2J}/*ob*^{2J} strain that is associated with an absence of RNA. The data showing that one strain overproduces *ob* RNA while the other fails to express this gene, make it unlikely that these changes are secondary. This conclusion can be confirmed using transgenic mice to complement the obese mutation. The nature of the mutation in the co-isogenic SM/Ckc- + ^{Dac}*ob*^{2J}/*ob*^{2J} mice is as yet unknown. Preliminary analysis of the gene structure has indicated

that the polymorphic *Bgl*II site maps ~7 kb upstream of the promoter. As this mutation is co-isogenic, the data strongly suggest that this polymorphism is associated with the mutation, although it is not yet clear whether there is a structural alteration (insertion or deletion) or base change. Further studies will be required to identify the molecular basis for the absence of *ob* RNA expression in SM/Ckc- + ^{Dac}*ob*^{2J}/*ob*^{2J} mice, as well as the molecular mechanisms that result in the increased expression of *ob* RNA in C57BL/6J *ob/ob* mice.

The increased level of *ob* RNA in the C57BL/6J mutant, which has greatly increased fat cell mass, raises the possibility that the level of expression of this gene signals the size of the adipose depot. This hypothesis is consistent with the lipostasis theory, implying that an increase in the level of the *ob* signal (as might occur after a prolonged binge of eating) may act directly or indirectly on the CNS to inhibit food intake and/or regulate energy expenditure as part of a homeostatic mechanism to maintain constancy of the adipose mass. Such effects could be explained if the *ob* protein activated the sympathetic nervous system via interactions with the VMH. Direct effects of *ob* on the CNS would require that mechanisms exist to allow its passage across the blood-brain barrier. The means by which the CNS effects changes in body weight remain to be elucidated and may include mechanisms independent of effects on food intake and energy expenditure. In addition, *ob* may have endocrine, paracrine and autocrine effects on different tissues. These and other possibilities will be testable if the active form of the *ob* protein can be shown to circulate in plasma.

The *ob* mutation is associated with a myriad of hormonal and metabolic alterations³⁵. These include effects on thermoregulation, fertility, adrenal and thyroid function as well as a wide range of biochemical abnormalities. Although we cannot exclude the possibility that many of these pleiotypic effects are the result of the expression of *ob* in sites not included in our initial survey of mouse tissues, the apparent expression of *ob* in only adipose tissue suggests that many of these changes are secondary. It has been suggested that the complex *ob* phenotype reflects an imbalance in the activity of the autonomic nervous system with low sympathetic and high parasympathetic tone³⁵. If *ob* is a signal molecule that regulates body weight, it may act by interacting with CNS (and other) receptors to modulate food intake and the activity of the autonomic nervous system. Parabiosis experiments suggest that the *ob* receptor is encoded by the mouse *db* (diabetes) gene⁶, a possibility that can be tested by positional cloning of *db* or by expression cloning and mapping of the *ob* receptor³⁶.

The extensive homology of the *ob* gene product among vertebrates suggests that its function is highly conserved. Now that the human homologue of *ob* is cloned, it will be possible to test for mutations in the human *ob* gene. As *ob* heterozygotes have an enhanced ability to survive a prolonged fast²³, heterozygous mutations at *ob* may provide a selective advantage in human populations subjected to caloric deprivation³⁷. Identification of *ob* now offers an entry point into the pathways that regulate adiposity and body weight and should provide a fuller understanding of the pathogenesis of obesity. □

Received 17 October; accepted 4 November 1994.

- Burros, M. Despite awareness of risks, more in US are getting fat in *The New York Times* 17 July 1994.
- Grund, S. M. & Barnett, J. P. *Disease-a-Month* **36**, 645–696 (1990).
- Friedman, J. M. & Leibel, R. L. *Cell* **69**, 217–220 (1990).
- Ingalls, A. M., Dickie, M. M. & Snell, G. D. *J. Hered.* **41**, 317–318 (1950).
- Friedman, J. M. et al. *Genomics* **11**, 1054–1062 (1991).
- Coleman, D. L. *Diabetologia* **14**, 141–148 (1978).
- Dickie, M. M. & Lane, P. W. *Mouse News Lett.* **17**, 52 (1957).
- Bahary, N. et al. *Mamm. Genome* **4**, 511–515 (1993).
- Walther, C. et al. *Genomics* **11**, 424–434 (1991).
- Riley, J. et al. *Nucleic Acids Res.* **18**, 2887–2890 (1990).
- Hermanson, G. G. et al. *Nucleic Acids Res.* **19**, 4943–4948 (1991).
- Sternberg, N. *Trends Genet.* **8**, 11–16 (1992).
- Hartl, D. L. & Nurminksky, D. I. *BioTechniques* **15**, 201–208 (1993).
- Dietrich, W. et al. *Genetics* **131**, 423–447 (1992).
- Church, D. M. et al. *Nature Genet.* **6**, 98–105 (1994).

- Gish, W. & States, D. J. *Nature Genet.* **3**, 266–272 (1993).
- Kozak, M. *Cell* **44**, 283–292 (1986).
- Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8998–9002 (1988).
- Padgett, R. A. et al. *Rev. Biochem.* **55**, 1119–1150 (1986).
- Cooper, D. N. & Youssoufian, H. *Hum. Genet.* **78**, 151–155 (1988).
- Heijne, G. v. *Nucleic Acids Res.* **14**, 4683–4690 (1986).
- Sabatini, D. D. & Adesnick, M. B. *The Metabolic Basis of Inherited Disease* (eds Scriver, C. R. et al.) 177–223 (McGraw-Hill, New York, 1989).
- Coleman, D. L. *Science* **203**, 663–665 (1979).
- Bray, G. A. *Int. J. Obesity* **14**, 909–926 (1990).
- Rubner, M. *Ztschr. Biol.* **30**, 73–142 (1894).
- Adolph, E. F. *Am. J. Physiol.* **151**, 110–125 (1947).
- Hervey, G. R. *Nature* **222**, 629–631 (1969).
- Cannon, W. B. & Washburn, A. L. *Am. J. Physiol.* **29**, 441–454 (1912).
- Brobeck, J. R. *Physiol. Rev.* **25**, 541–559 (1946).
- Hetherington, A. W. & Ranson, S. W. *Am. J. Physiol.* **136**, 609–617 (1942).
- Brobeck, J. R. *Yale J. Biol. Med.* **20**, 545–552 (1948).

32. Kennedy, G. C. *Proc. R. Soc. B* **140**, 578–592 (1953).
33. Mayer, J. *Ann. N.Y. Acad. Sci.* **63**, 15–43 (1955).
34. Hervey, G. R. *J. Physiol.* **145**, 336–352 (1959).
35. Bray, G. A. *J. Nutrition* **121**, 1146–1162 (1991).
36. Bahary, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8642–8646 (1990).
37. Neel, J. V. *Am. J. hum. Genet.* **14**, 353–362 (1962).
38. Larin, Z., Monaco, A. P. & Lehrach, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4123–4127 (1991).
39. Green, E. D. & Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1213–1217 (1990).
40. Kusumi, K. et al. *Mamm. Genome* **4**, 391–392 (1993).
41. Vidal, S. M. et al. *Cell* **73**, 469–485 (1993).
42. Friedman, J. M. et al. *Genomics* **11**, 1054–1062 (1991).
43. Chirgwin, J. M. et al. *Biochemistry* **18**, 5294–5299 (1979).
44. Friedman, J. M., Cheng, E. & Darnell, J. E. *J. molec. Biol.* **179**, 37–53 (1984).
45. Wang, A. M., Doyle, M. V. & Mark, D. F. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9717–9721 (1989).
46. Benton, W. D. & Davis, R. W. *Science* **195**, 180–182 (1977).
47. Burmeister, M. & Lehrach, H. *Trends. Genet.* **5**, 41 (1989).
48. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
49. Venter, C. J. *Automated DNA Sequencing and Analysis* (Academic, New York, 1993).
50. Yu, Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9931–9935 (1989).

ACKNOWLEDGEMENTS. We thank D. Coleman for prior genetic and physiological analysis; R. Leibel and N. Bahary for their important contributions to the early phases of this work; D. Koos and R. Cox for help in isolating the first set of YAC clones; A. Buckler, J. Rutter and C. Stotter for instruction in exon trapping; Y. Yu and B. Luring for advice on *in vitro* translation and for canine pancreatic microsomal membranes; P. Gruss for providing the *Pax4* probe before publication and communicating its map position; J. E. Darnell, N. Heintz, R. Kucherlapati, S. Burley, J. Froude, D. Luck, C. Winestock and L. Safani for comments; S. Korres for preparing the manuscript; J. Sholtis for photography; F. Ilchert for help in preparing the figures; A. Popowicz, P. Dash and the staff of computing services for assistance; D. Sabatini for expert advice on the characterization of the *ob* polypeptide; and J. E. Darnell Jr, who provided the environment to initiate this work. J.M.F. is an investigator of the Howard Hughes Medical Institute. This work was supported by a grant from NIH/NIDDK.

Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin

Chris Oubridge, Nobutoshi Ito*, Philip R. Evans, C.-Hiang Teo* & Kiyoshi Nagai

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The crystal structure of the RNA-binding domain of the small nuclear ribonucleoprotein U1A bound to a 21-nucleotide RNA hairpin has been determined at 1.92 Å resolution. The ten-nucleotide RNA loop binds to the surface of the β -sheet as an open structure, and the AUUGCAC sequence of the loop interacts extensively with the conserved RNP1 and RNP2 motifs and the C-terminal extension of the RNP domain. These interactions include stacking of RNA bases with aromatic side chains of proteins and many direct and water-mediated hydrogen bonds. The structure reveals the stereochemical basis for sequence-specific RNA recognition by the RNP domain.

THE U1A protein is a component of the U1 small nuclear ribonucleoprotein particle (U1 snRNP), which is one of five large RNA-protein complexes involved in pre-mRNA splicing^{1,3}. U1A protein contains two copies of the RNP motif separated by a protease-sensitive linker⁴. This motif functions as an RNA-binding module in over 150 distinct proteins involved in RNA processing and transport^{5,6}. The RNP motif is characterized by two short highly conserved sequences, called RNP1 and RNP2, embedded within a more weakly conserved sequence of ~80 amino acid residues. Proteins containing the RNP motif are known to bind various forms of RNA, including single-strands, hairpins and internal loops. U1A protein binds to hairpin II of U1 snRNA^{7,8} (Fig. 1c), and a fragment containing the first 98 residues retains the full RNA-binding properties of the whole protein (Fig. 1a)^{9,14}. A crystallographic study of a U1A protein fragment has shown that the region between residues 5 to 85 forms a compact domain (hereafter called the RNP domain) consisting of a four-stranded β -sheet flanked on one side by two α -helices¹⁰. The RNP1 and RNP2 motifs lie side-by-side on the two central β -strands (β_3 and β_1) of the β -sheet¹⁰. NMR studies of fragments of U1A protein confirmed that the structure was

similar in solution^{15,16}. RNA binding studies of various point mutants of U1A protein showed that residues within the RNP1 and RNP2 motifs are likely to be involved in RNA binding, and that basic residues on the loops between the β_1 strand and helix(A) and between the β_2 and β_3 strands also make important interactions^{10,11}.

U1A protein is closely related to a protein associated with the U2 snRNP called U2B' protein¹⁷ (Fig. 1a), which binds to hairpin IV of U2 snRNA only when it is complexed with U2A' protein (unrelated to U1A protein). The RNA-binding sites for U1A protein and the U2B'–U2A' protein complex share the AUUGCA hexanucleotide sequence in the loop (Fig. 1c, d), but these proteins bind specifically to their respective cognate RNA hairpins^{9,12}. RNA-binding studies of various chimaeras of U1A and U2B' proteins have identified residues necessary for the interaction between U2A' and U2B' proteins and those residues that discriminate between the two cognate RNAs^{9,12}. The results of mutagenesis experiments^{9,12} combined with ethylation protection¹¹ of U1 RNA hairpin II have enabled us to propose a model of the complex between the U1A protein and hairpin II of U1 snRNA¹¹. In this model, the RNA interacts with the surface of the four-stranded β -sheet. NMR studies of the interactions between the RNP domain of heterogenous (hn) RNP protein C and r(U)₈ (refs 18, 19) and between the N-terminal

* Present addresses: Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK (N.I.); EMBL, Meyerhofstrasse 1, Heidelberg D-69117, Germany (C.-H.T.).

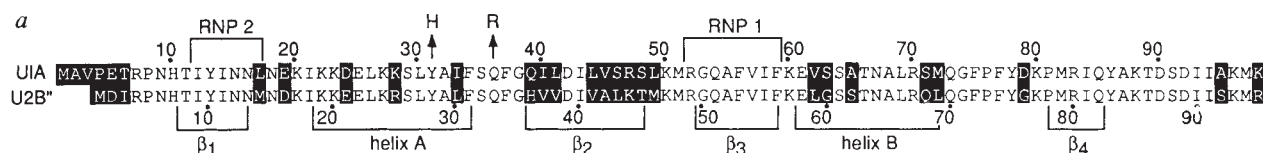


FIG. 1. *a*, Amino-acid sequence of the RNA-binding domain of the U1A and U2B' proteins (single-letter code)^{4,17}. Positions where the two proteins have different amino acids are highlighted. The two mutations (Tyr 31→His and Gln 36→Arg) of the construct used for co-crystallization are indicated with arrows. *b*, The synthetic 21-nucleotide RNA sequence used for co-crystallization. The 7 nucleotides of the loop that form close contacts with U1A protein are highlighted. *c*, U1A protein-binding site (hairpin II) of U1 snRNA. The numbering starts from the trimethyl guanidine cap as the first nucleotide. *d*, Part of hairpin IV of U2 snRNA, the binding site for the U2B' protein–U2A' protein complex. The highlighted region indicates the 6 nucleotides of the loop that are the same in U1 snRNA stem/loop II. The boxed G is an important determinant of specificity⁹. *e*, Predicted folding of the two U1A protein-binding sites in the 3'UTR of U1A pre-mRNA, with the 7-nucleotide internal loop sequences highlighted²¹.

METHODS. Variants of the U1A protein were overproduced as described¹¹. RNA oligonucleotides were synthesized on an ABI 392 DNA/RNA synthesizer using phosphoramidite chemistry with *t*-butyldimethylsilyl protecting groups²². Oligonucleotides were deprotected⁵⁰, gel-purified and extracted from gel pieces by electroelution. Initial crystallizations with U1A protein fragments having the wild-type sequence persistently resulted in poorly diffracting cubic crystals. As the free U1A protein crystal also had a cubic space group¹⁰, it was thought that the

fragment of U1A protein and U1 snRNA hairpin II (ref. 15) are in good agreement with our model.

It has been found that U1A protein binds to the 3'-untranslated region (3'UTR) of its own pre-mRNA²⁰. Two molecules of U1A protein bind to a region upstream of the polyadenylation signal and inhibit cleavage and polyadenylation of the pre-mRNA, thus autoregulating its expression. The binding site has an evolutionarily conserved secondary structure with two internal loops, one containing the AUUGCAC sequence found in U1 RNA hairpin II and the other with a variant AUUGUAC sequence (Fig. 1e)²¹.

Here we describe the crystal structure of the complex between the N-terminal RNP domain of U1A protein and a 21-nucleotide RNA representing hairpin II of U1 snRNA. This structure reveals the precise stereochemistry of the sequence-specific interaction between U1A protein and the U1 RNA hairpin, and it also suggests how the RNP domains of other proteins binds to their cognate RNAs.

Structure determination

An extensive crystallization search was carried out with various N-terminal fragments of U1A protein¹¹ and synthetic RNA²². Initial crystallization with the wild-type sequence resulted in poorly diffracting cubic crystals, so variant proteins with surface mutations were included in the crystallization trials. The best crystals, which diffract to 1.7 Å resolution, were obtained with a 21-nucleotide RNA, r(AAUCCAUGGCACUCCGGAUU), and a protein containing residues between 2 and 98 with two mutations, Tyr 31→His and Gln 36→Arg (Fig. 1*a*). These two mutations are located away from the RNA-binding surface, as characterized biochemically¹¹, and the mutant shows RNA-binding properties indistinguishable from those of a wild-type protein of the same length (data not shown). The nucleotide sequence of the 21-nucleotide RNA is authentic, except for the last base pair and the overhanging nucleotide (Fig. 1*b, c*):

same protein–protein contacts might be used to pack both crystals. Various mutations were introduced to surface residues to disrupt such interactions. Crystallization trials with the double mutant (Tyr 31→His, Gln 36→Arg) and a 21-nucleotide RNA resulted in crystals of a hexagonal space group with excellent diffraction qualities. Those were grown at 20 °C from 1.8 M ammonium sulphate, 40 mM Tris-HCl, pH 7.0, and 5 mM spermine, using the hanging-drop vapour-diffusion method.

nucleotide numbers given in the text are as shown in Fig. 1*b*. Crystals are hexagonal prisms, typically growing to a length of 600 μm and a width of 300 μm . The space group is $P6_522$, with unit cell edges of $a=b=97.0$ Å, $c=255.3$ Å. The structure was solved using a native dataset and datasets at two different wavelengths from a methyl mercury derivative of a Ser 29→Cys mutant²³; all data were collected at a temperature of 100 K (Table 1).

Figure 2*a* shows a portion of the experimental electron density map at 2.4 Å resolution with isomorphous replacement phases after solvent-flattening. This map is of excellent quality, and the initial atomic model was readily built into it apart from U20 and overhanging U21 at the 3' end of the RNA stem, three nucleotides in the loop (U13, C14 and C15) and a few amino acids at the N- and C-termini. Analysis of the RNA extracted from the crystal showed that the RNA loop is intact and the discontinuity of the RNA density is due to disorder (data not shown). The structure has been refined using the TNT program²⁴ to 1.92 Å resolution: the current crystallographic *R*-factor is 21.5%, with 410 water molecules in the model (Table 1). Figure 2*b* shows a region of the refined $2F_{\text{obs}} - F_{\text{calc}}$ map at 1.92 Å resolution. Well-ordered water molecules are found bound to 2'-OH and phosphate groups on the RNA molecules as well as at the protein-RNA interface and protein surface. The asymmetric unit contains three copies of the complex which are related by a non-crystallographic three-fold axis. The structures of the complexes are essentially identical except in disordered parts of the molecule. Regions ordered in the protein crystal (residues 5-47 and 51-85; ref. 10) undergo no significant structural changes upon RNA binding except for residues 79 and 80. The residues at the N- and C-termini and of the flexible $\beta 2$ - $\beta 3$ loop (residues 48-50) have become ordered and a short α -helix centred around residue 95 is present in the complex crystal (Fig. 3). This helix was first observed by Howe *et al.*¹⁵ in an NMR study of the free protein.

TABLE 1 X-ray structure determination

Data set	Native	Derivative I CH ₃ Hg (S29C)	Derivative II CH ₃ Hg (S29C)
Resolution range	33.5–1.92 Å	22.9–2.42 Å	79–2.3 Å
Wavelength	0.882 Å	0.882 Å	1.000 Å
Number of unique reflections	54,400	27,595	24,314
Mean redundancy	7.7	4.8	13.4
Completeness	99.7%	99.2%	75.0%
Overall R_{sym} *	0.038	0.025	0.098
R_{diff} †	—	0.196	0.211
Phasing:			
Number of sites	—	3	3
Phasing power‡			
Acentric	—	1.8	1.8
Centric	—	1.3	1.3
Cullis R -factor§	—	0.60	0.61
Mean FOM			
Overall	0.6876		
Acentric	0.6789		
Centric	0.7266		
Solvent flattening:			
No. of cycles	10		
Resolution range	17.9–2.4 Å		
Mean overall FOM	0.8548		

Data collection. Native and derivative I data sets were collected at the SERC Daresbury Laboratory on synchrotron beamline PX9.6 using a 30-cm MAR IP scanner. The derivative II dataset was collected at the KEK Photon Factory on synchrotron beamline BL-6A using 40×20 cm Fuji IPs on a Weissenberg camera³⁹. Data from the Photon Factory were reduced using DENZO (Z. Otwinowski, personal communication); data from Daresbury were reduced with MOSFLM⁴⁰. Scaling and processing was done using the CCP4 suite of programs⁴¹. Each dataset was collected from a single crystal cooled to 100 K with an Oxford Cryosystems 600 Series Cryostream Cooler⁴². Crystals were transferred from mother liquor to cryoprotectant solution (1.9 M ammonium sulphate, 25% glycerol, 5 mM spermine, 40 mM Tris-HCl, pH 7.0) 15 min before freezing, mounted in a rayon loop⁴⁵ and frozen directly in the cold gas stream. **Phasing.** Two out of three heavy-atom sites were initially identified with the SHELXS-90 program⁴⁴ in Patterson search mode; the third site was found in a difference Fourier map using phases calculated and refined from the initial two sites with MLPHARE⁴⁵. Two of the sites were found to have double conformations and these were treated as 'split' sites. All five heavy-atom positions were refined and phases calculated to 2.4 Å with MLPHARE, including data from derivatives at two wavelengths (0.882 Å and 1.000 Å). An electron density map was calculated and then subjected to 10 rounds of solvent flattening using the SOLOMON program⁴⁶ to 2.4 Å resolution. **Model building.** Model building was carried out on the solvent-flattened map in the graphics program O⁴⁷. The protein molecules were built mostly by fitting coordinates from the uncomplexed protein structure and moving atoms to fit better into the electron density map with the O build options. The initial RNA model was based on the co-ordinates of the proposed structure of Krol *et al.*⁴⁸ and fitted into the density in the same manner as the protein. **Refinement.** The model was refined against the native data to 1.92 Å using the TNT program²⁴. The initial model had an R -factor of 42.7% against all the data between 10 and 1.92 Å. The refinement has produced a model with an R -factor of 21.5% with restrained atomic B -factors, and a free R -factor⁴⁹ of 26.6%. The structure has a bond length r.m.s. deviation of 0.015 Å and a bond angle r.m.s. deviation of 2.12°.

* $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$, the average intensity from multiple measurements.

† $R_{\text{diff}} = \sum ||F_{\text{PH}}| - |F_{\text{P}}|| / \sum |F_{\text{P}}|$.

‡ Phasing power is r.m.s. $(|F_{\text{h}}|/E)$, where $|F_{\text{h}}|$ is the heavy-atom structure factor amplitude and E the residual lack of closure error.

§ Cullis R -factor, $(\sum |E| / \sum ||F_{\text{PH}}| - |F_{\text{P}}||)$.

|| FOM, figure of merit.

Overall structure of the complex

The overall structure of one of the three complexes in the asymmetric unit is shown in Fig. 4a and b. All the structural features described here are conserved in the three complexes unless stated otherwise. The ten-nucleotide loop of the RNA hairpin is bound to the protein as an open structure with most of the bases splayed

out from its centre. The RNA loop extends from the double-helical stem in a direction nearly perpendicular to the helix axis and interacts with a large area of the four-stranded β -sheet (Fig. 4a). The positioning of RNA with respect to the protein is similar to that predicted by Jessen *et al.*¹¹, but their interaction is significantly different in detail. The $\beta 2$ – $\beta 3$ protein loop protrudes through the RNA loop where it forks from the stem and prevents the pairing of bases within the RNA loop (Fig. 4b). The sugar–phosphate backbone of the 5' stem interacts with the Lys 20 and Lys 22 side chains located on the loop between $\beta 1$ and helix-A. Howe *et al.*¹⁵ observed that the imino proton resonances of U60 and U61 shift upon U1A protein binding. U61 corresponds to A1 in the crystal (Fig. 1b and c), but there is no corresponding nucleotide to U60 in the shorter RNA used for crystallization. The imino proton of U61 shows nuclear Overhauser effects (NOE) from contacts to a protein side chain. This residue is likely to be Lys 22 but it lies outside NOE distance from A1 in the crystal structure (Fig. 4a). Lysine and arginine residues previously shown to be important for RNA binding (Fig. 4 of ref. 10) follow the phosphate backbone of the bound RNA and show the importance of electrostatic interactions in positioning the RNA loop. The AUUGCAC sequence in the RNA loop fits into the groove between the $\beta 2$ – $\beta 3$ protein loop and the C-terminal region (Fig. 4b) and interacts as a single-stranded RNA with the side chains of the conserved RNPI and RNP2 motifs and the main-chain amide and carbonyl groups of the C-terminal region.

The last three nucleotides in the loop, UCC, extend into solution and form no apparent contacts with the β -sheet (Fig. 4a). These nucleotides in one of the complexes interact with equivalent nucleotides of another complex related by a crystallographic dyad. The same three nucleotides in the other two complexes interact with each other in a similar manner around a non-crystallographic dyad. The electron density is unambiguous up to the 5' phosphate of C13 and beyond the 5' phosphate of G16, but is not clear between these two phosphates. The structure of these three nucleotides is likely to be affected by crystal contacts, but biochemical evidence suggests that they are mobile and not in contact with the protein in solution; *in vitro* selection with randomized 10-nucleotide loops yielded RNA molecules with the AUUGCAC sequence followed by random three nucleotides²⁵. Our ethylation protection experiment showed that the phosphate backbone in this region is pointing towards the protein¹¹; these three nucleotides probably act merely as a linker between the AUUGCAC sequence and the stem. The stacking interaction between adjacent bases and the helical conformation of the sugar–phosphate backbone extend from the 5' stem into the first two unpaired nucleotides of the loop, A6 and U7. The backbone conformation of the remainder of the loop is irregular; the sugar conformation of U8 and C10 is 2'-endo, whereas all other ordered sugars show 3'-endo conformation.

Hydrogen-bonding network

The C5–G16 base pair that closes the RNA loop plays a key part in the positioning of the RNA. Arg 52, one of the RNPI residues, hydrogen-bonds with N7 and O6 of G16 and with N1 of A6 (Fig. 5a). This is consistent with the shift of imino proton resonance on G16 (G76 in ref. 15). The Arg 52 side chain is buttressed by hydrogen bonds between its N ϵ H and the main-chain carbonyl groups of Arg 47 and Ser 48 and between Arg 52 N η H and the carbonyl group of Ser 48. The main-chain amide group of Arg 52 is also hydrogen-bonded to the main-chain carbonyl group of Arg 47. The amide group of Arg 47 forms a hydrogen bond with the phosphate group of G16. The polypeptide loop between residues Arg 47 and Met 51 ($\beta 2$ – $\beta 3$ loop) is flexible in crystals of the free protein¹⁰ but has become fixed upon RNA binding as part of the extensive hydrogen bond network with RNA. The high degree of structure of the $\beta 2$ – $\beta 3$ loop has also a more general and profound influence on RNA binding as it protrudes through the RNA loop and opens it, thereby

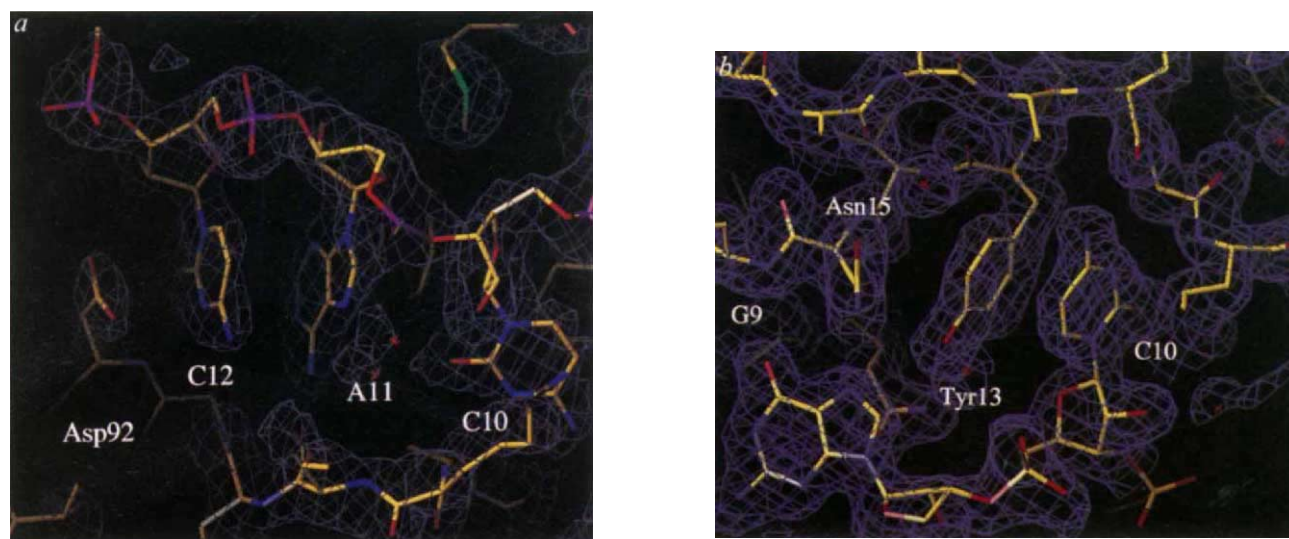
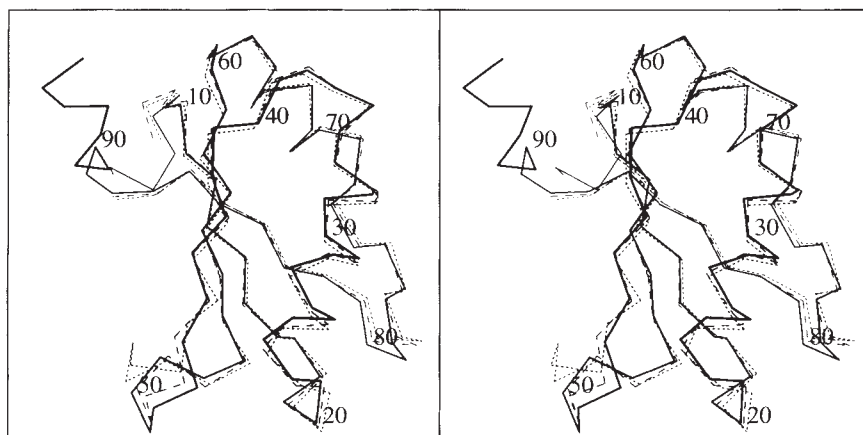


FIG. 2 *a*, Experimental electron density map at 2.4 Å resolution calculated with isomorphous replacement phases after solvent flattening (contoured at 1σ) which is superimposed on the refined atomic model

(1.92 Å resolution; *R*-factor, 21.5%). *b*, Portion of the refined ($2F_{\text{obs}} - F_{\text{calc}}$) map at 1.92 Å resolution contoured at 1σ , where 'obs' and 'calc' denote observed and calculated, respectively.

FIG. 3 Comparison of the α -carbon positions in a crystal of the free U1A protein crystal and a crystal of the RNA complex. α -Carbon positions of a pair of protein molecules are superimposed using the least-square fitting option of the O program⁴⁷. U1A protein in the RNA complex crystal is shown as a solid line; the structure of the two molecules in the protein crystal¹⁰ are shown as dashed and dotted lines. The most significant differences between free protein and complex crystals are seen in the $\beta 2$ - $\beta 3$ loop (48-50), the helix(B)- $\beta 4$ loop (79-80) and the appearance of clear electron density for residues 86 to 98 at the C terminus.



presenting the loop nucleotides to the protein as a single-stranded nucleic acid. Replacement of Arg 52 with Gln completely abolishes RNA binding. This is partially restored in the Arg 52→Lys mutant^{10,11}.

The pyrimidine ring of U7 stacks on the adenine ring of A6 (Fig. 5*a*). O2 and N3H of U7 are hydrogen-bonded to the 2-amino group of G9 and the carboxyl of Glu 19 respectively (Fig. 5*b*). The O6 of G9 forms a hydrogen bond to the main-chain amide group of Asn 16 and a water-mediated hydrogen bond to the main-chain carbonyl group of Leu 17. The N7 of G9 is hydrogen-bonded to the side chain of Asn 15 and the N3 of G9 forms a water-mediated hydrogen bond to the 2'-OH group of the U7 ribose as well as the main-chain carbonyl group of Leu 49.

The pyrimidine ring of U8 lies above the bases of U7 and G9 and partially stacks to the purine ring of G9, which in turn stacks to the side chain of Gln 54 (Fig. 4*a*). The O2 and O4 of U8 form water-mediated and direct hydrogen bonds to the side chains of Arg 83 and Lys 80, respectively (Fig. 5*c*). The N3H of U8 is hydrogen-bonded to the side chain of Asn 16, which is in turn hydrogen-bonded to the main-chain carbonyl of Pro 81. These hydrogen bonds are probably responsible for the movement of Asp 79 and Lys 80 upon RNA binding (Fig. 3*b*).

The pyrimidine ring of C10 stacks on the phenyl ring of Tyr 13 (Fig. 5*d*). The 4-amino group, N3H and 2-keto-oxygen of C10 form hydrogen bonds to the side chain of Gln 85, the main-chain carbonyl group of Tyr 86 and the main-chain amide group of Lys 88, respectively (Fig. 5*d*). The phenyl ring of Tyr 13 is held in place by a hydrogen bond to the side chain of Gln 54 which is in turn hydrogen bonded to the main-chain carbonyl groups of Lys 50 and Arg 52. A Tyr 13→Phe replacement abolished RNA binding¹¹. As shown in Fig. 5*a*, the main chain of this region is involved in an extensive hydrogen-bond network. The loss of a hydrogen bond with Gln 54 would fail to hold the phenyl ring in a position favourable for stacking with the C10 ring and also affect the hydrogen-bond formation between the Gln 54 side chain and the main-chain carbonyl groups. Phe or Tyr are often found in corresponding positions in other RNP domains, but the replacement of Gln 54 with Phe also severely reduces the RNA-binding affinity.

Bases of A11 and C12 are sandwiched between Phe 56 and Asp 92, forming a four-element stack (Fig. 5*e*). As shown in Fig. 5*e*, nearly all the hydrogen-bond donors and acceptors of A11 and C12 are involved in hydrogen bonding, mainly with main-chain amide and carbonyl groups of the C-terminal region.

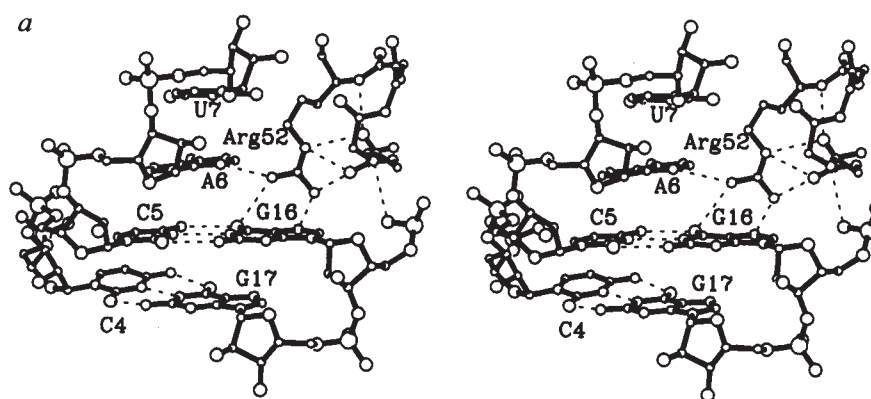
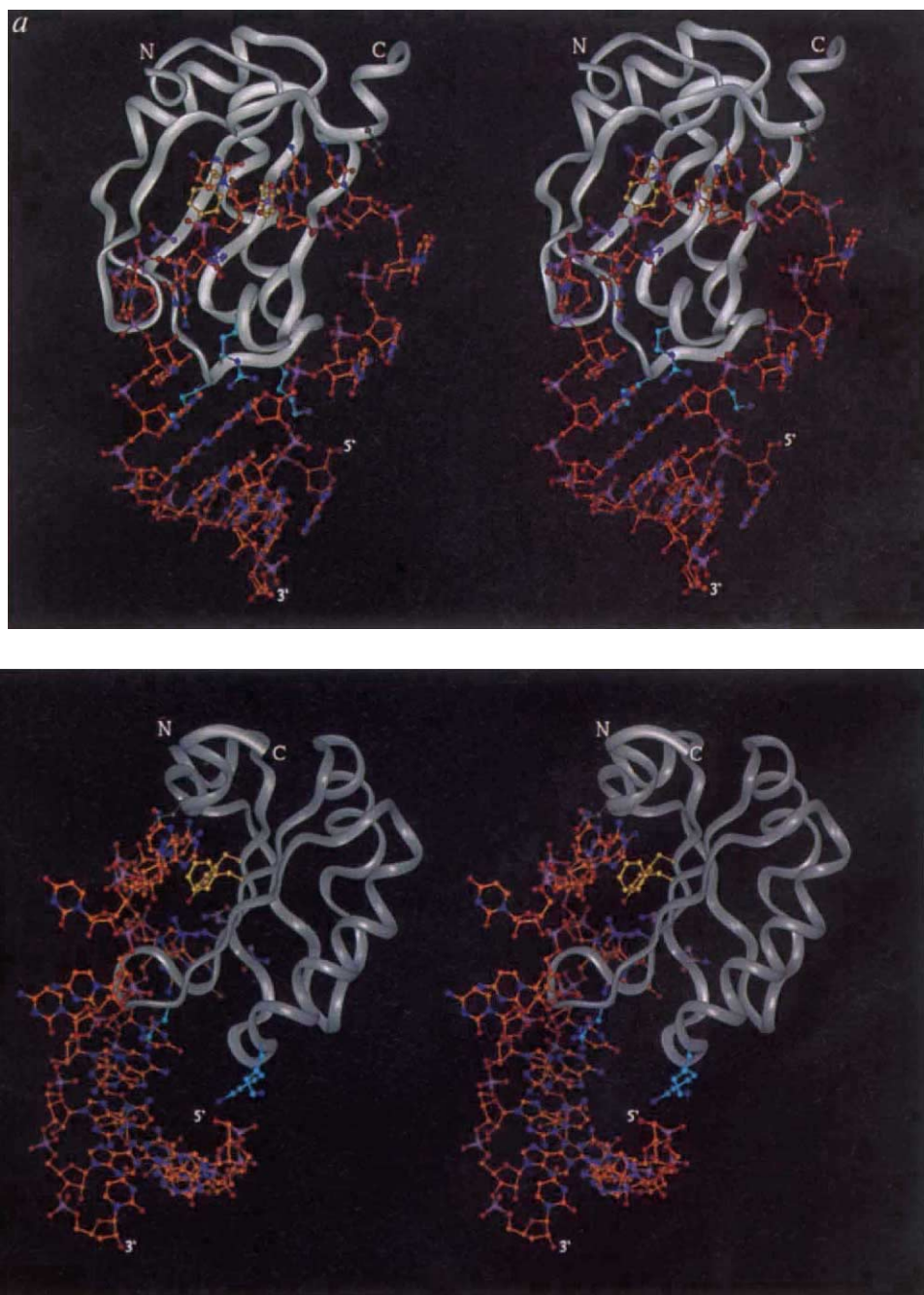


FIG. 4 Two views of the structure of one complex in the asymmetric unit. The protein main chain between residues 3 and 98 is represented as a white ribbon. The side chains of surface residues important for RNA binding are included and N and O atoms are coloured blue and red, respectively. Carbon atoms of these residues are coloured yellow (Tyr 13 and Phe 56), pink (Asn 15 and Asn 16), purple (Gln 54), light blue (Lys 20, Lys 22 and Arg 52), and grey (Asp 92). Nucleotides 1 to 19 (see also Fig. 1c) are included and C, N, O and P atoms are orange, blue, red and purple, respectively. Nucleotides U13, C14 and C15 are not well represented by the density. These three nucleotides are not in contact with the protein; they are at a crystal contact and their positioning in the model is inaccurate and may be affected by crystal packing. *a*, RNA hairpin bound to the surface of the β -sheet with the loop shown face-on. The β -strands are numbered from the N terminus and RNP2 (containing Tyr 13, Asn 15 and Asn 16) and RNP1 (containing Arg 52, Gln 54 and Phe 56) motifs are on the β 1 and β 3 strands. *b*, View of complex, illustrating right-handed twist of antiparallel β -sheet and the groove formed between the β 2- β 3 loop, the β -sheet and the C-terminal region. The contacts between the stem and lysines 20 and 22 are also shown. The β 2- β 3 loop can be seen protruding through the loop of the RNA to stabilize the open structure of the RNA loop.



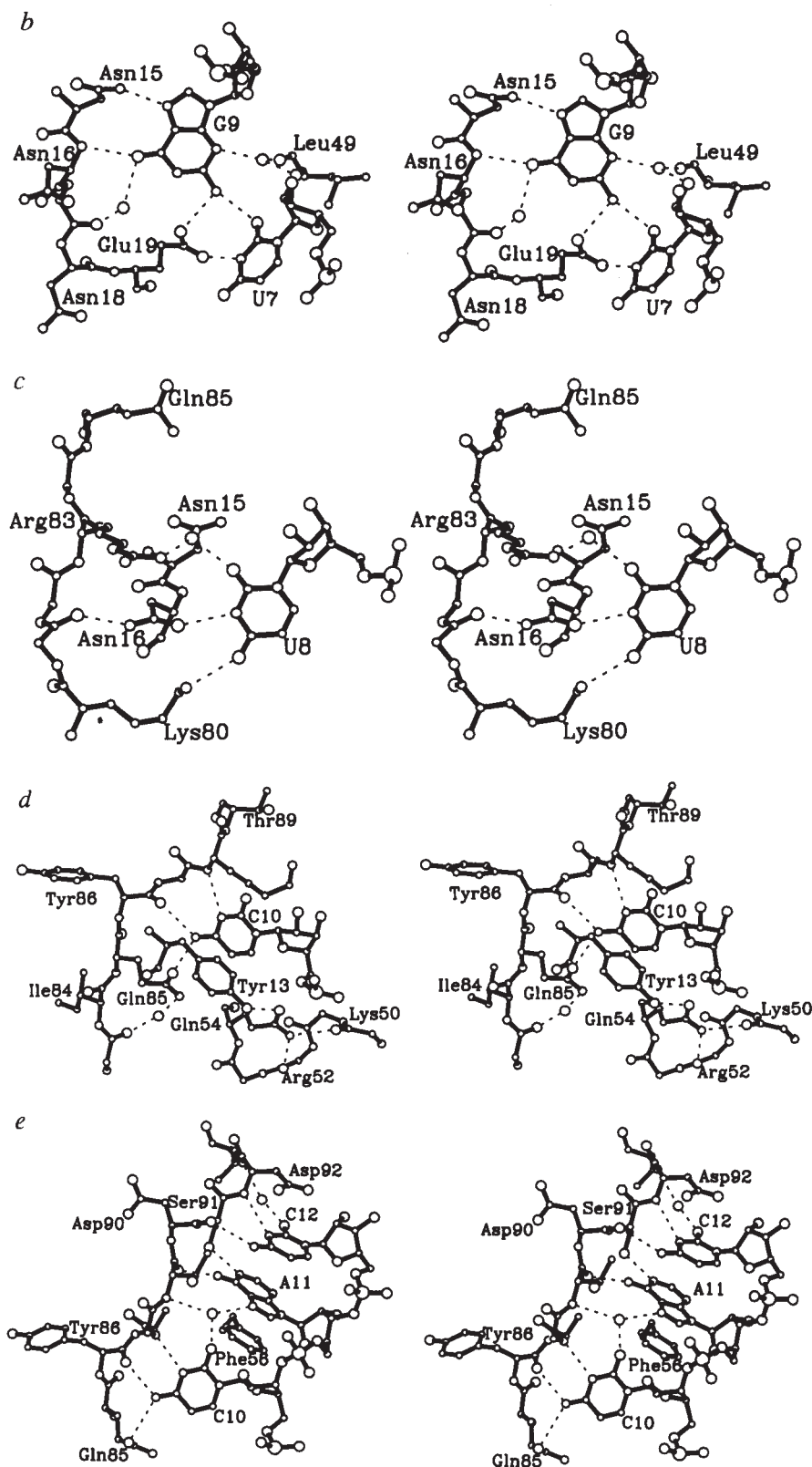


FIG. 5 The interaction of the AUUGCAC heptanucleotide with the U1A protein. *a*, The fork where the stem of the hairpin ends and the loop begins, including two base pairs in the stem and nucleotides A6 and U7 which continue the stacking of the stem into the loop region. The critical residue Arg 52 is shown hydrogen-bonding to A6 and G16, and is buttressed by main-chain carbonyl oxygens. The amide group of Arg 47 also makes a hydrogen bond with the phosphate of G16. *b*, Contacts between nucleotides U7 and G9 and amino acids in the conserved RNP1 and RNP2 regions. The stem stacking does not continue past U7, which makes a hydrogen-bond contact with G9. The carboxyl group of Glu 19 makes contacts with both U7 and G9 and Asn 15 makes a hydrogen bond to G9. G9 is also contacted by a main-chain amide and carbonyl from this region. *c*, Nucleotide U8 makes contacts with Asn 16 of RNP2 and with Lys 80 and Arg 83 (via H_2O) residues in the $\beta 4$ strand. *d*, The pyrimidine ring of C10 stacks on Tyr 13 and is also hydrogen-bonded to main-chain groups of the C terminus and the side chain of Gln 55. The buttressing of Tyr 13 by a hydrogen bond between its phenolic oxygen and the Gln 54 side chain is a requirement for RNA binding. *e*, The A11 and C12 bases are stacked between the Phe 56 ring and the carboxyl group of Asp 92. The phosphate backbone follows the main chain of the C terminus, and main-chain carbonyls and amides make numerous interactions with C10, A11 and C12.

Role of the C-terminal region

The N- and C-terminal polypeptides extending from the $\beta 1$ and $\beta 4$ strands interact with each other through direct and water-mediated hydrogen bonds. The C-terminal polypeptide extends across the β -sheet and forms a short α -helix (helix-C). The main-chain amide and carbonyl groups in the C-terminal region participate in base recognition by forming direct and water-mediated hydrogen bonds (Fig. 5e). The structure of the C-terminal region

is stabilized by hydrophobic interactions between Ile 58 (RNP1 residue), Ile 93 and Ile 94. Replacement of Thr 11 by Val completely abolishes RNA binding¹¹, and we had proposed that this residue might interact directly with RNA. In fact, Thr 11 is now seen to be crucial in holding the C-terminal region in place by forming a hydrogen bond with Ser 91 (not shown). A truncated protein containing residues 2–95 does not bind RNA, but addition of Lys 96 is sufficient to allow this protein to bind RNA¹⁰. Lys 96 and Lys 98 make no direct contacts with RNA nor with other parts of the protein, but the positive charges of these

lysines apparently play a key role in positioning the RNA. The C-terminal extension of the RNP domain has similarly been shown to be important for RNA binding in hnRNP C protein²⁶ and U1 70K protein²⁷.

Discussion

The structure of the complex described immediately suggests how U1A protein binds to the 3' UTR of its pre-mRNA²⁰. U1A

protein binds to a double-helical region with two internal loops located upstream of the polyadenylation signal (Fig. 1e). The two internal loops contain the AUUGCAC sequence and its variant AUUGUAC; both are also closed by a C·G basepair²¹ (C12–G51 and C25–G38 in Fig. 1e). Hence all the RNA–protein contacts shown in Fig. 5a–e are likely to be conserved in the 3'UTR complex. The last three nucleotides, UCC, make no protein contacts in the U1 RNA hairpin II complex and can be replaced with the base pair and single nucleotide found in the corresponding position in the two RNA internal loops.

The AUUGCA hexanucleotide is also conserved in U2 RNA hairpin IV, the binding site for the closely related U2B' protein (Fig. 1a). But there are three important differences in U2 RNA hairpin IV which could destabilize its binding to U1A protein (Fig. 1c, d): (1) the C·G base pair, which closes the loop and interacts with Arg 52, is replaced by a U·U base pair; (2) C in the seventh position in the loop of U1 RNA hairpin is replaced by G; (3) there are 11 nucleotides instead of 10 in the loop. This explains why U1A protein does not bind to U2 RNA hairpin IV (refs 9, 12). U2A' protein interacts with the β 2 strand and helix-A of U2B' protein^{9,11,12} and would provide additional interactions with the last four UACC nucleotides in the loop of U2 snRNA hairpin IV and stabilize its binding^{11,12}. We have obtained crystals of the ternary complex between U2B' protein, U2A' protein and hairpin IV of U2 RNA, which diffract to 3 Å resolution (S. R. Price, C.-H.T. and K.N., unpublished results) and its crystal structure will reveal the molecular details of this interaction.

In the U1A–U1 snRNA hairpin II complex, the AUUGCAC sequence is recognized by U1A protein as a single-stranded nucleic acid presented as part of a hairpin loop. NMR studies of the free U1 snRNA hairpin II by F. Allain and G. Varani (personal communication) have shown that the hairpin loop is largely unstructured in free RNA apart from the first three bases in the loop which stack with each other. The RNA loop therefore binds to U1A protein by an induced-fit mechanism. The binding of a single-stranded RNA containing the AUUGCAC sequence to U1A protein is substantially lower^{9,14} (T.-H. Jessen, K.N. and C.O., unpublished results). This implies that the stem is

important in restricting the conformation of the loop and reducing the entropy loss accompanying the protein binding, as well as allowing recognition of the terminal C·G base pair. The bases of these seven loop-nucleotides are splayed out, allowing nearly all the potential hydrogen-bond donors and acceptors of these bases to form direct or water-mediated hydrogen bonds with the protein. An unusually large number of these are to main-chain rather than side-chain groups on the protein. The bases of C10 and A11 show stacking interactions with the aromatic side chains of Tyr 13 and Phe 56. Such stacking interaction is not possible in double-helical DNA–protein complexes unless the DNA helix is severely distorted to disrupt the stacking of adjacent bases, as with TATA-box-binding protein^{28,29}.

In vitro selection experiments with other proteins of the RNP family showed that each RNP domain interacts with 5–7 nucleotides in a sequence-specific manner, but additional flanking nucleotides are necessary for high-affinity binding^{26,30,31}. The interaction between heptanucleotides and the RNP1 and RNP2 motifs is likely to be a conserved feature in all the RNP domains. However, the length of the protein loops and the distribution of basic amino acids are variable^{5,6}, so the binding of the flanking nucleotide sequence may be stabilized in a different manner for each RNP protein.

The crystal structures of three aminoacyl-tRNA synthetases complexed with their cognate tRNAs have been published^{32–34}. In the glutamyl- and aspartyl-tRNA synthetase complexes, the bases in the anticodon loop are splayed out and form many hydrogen bonds with the protein^{35,36}. In particular, the anticodon loop of the aspartyl-tRNA synthetase complex interacts as an extended single-stranded RNA with the surface of the β -barrel³⁶ in a manner similar to that seen in the complex described here (see Fig. 3 of ref. 36). In this complex the phenyl ring (Phe 127) on the surface of the β -sheet also stacks to the base of U35. In contrast, MS2 phage coat protein binds to a small RNA hairpin with rigid structure³⁷.

The β -sheet is a common structural element in many RNA-binding proteins^{16,38}, and the RNA–protein interactions observed in the U1A protein–RNA hairpin complex may be recurring features in many RNA–protein complexes. □

Received 15 September; accepted 24 October 1994.

- Lührmann, R., Kastner, B. & Bach, M. *Biochim. biophys. Acta* **1087**, 265–292 (1990).
- Steitz, J. A. et al. in *Structure and Function of Major and Minor snRNPs* (eds Birnstiel, M. L.) 115–154 (Springer, New York, 1988).
- Nagai, K. & Mattaj, I. W. in *RNA–protein interactions* (eds Nagai, K. & Mattaj, I. W.) 150–176 (IRL, Oxford, 1994).
- Sillekens, P. T. G., Habets, W. J., Beijer, R. P. & van Venrooij, W. J. *EMBO J.* **6**, 3841–3848 (1987).
- Burd, C. G. & Dreyfuss, G. *Science* **265**, 615–621 (1994).
- Birney, E., Kumar, S. & Krainer, A. *Nucleic Acids Res.* **21**, 5803–5816 (1993).
- Scherley, D. et al. *EMBO J.* **8**, 4163–4170 (1989).
- Bach, M., Krol, A. & Lührmann, R. *Nucleic Acids Res.* **18**, 449–457 (1990).
- Scherley, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. & Mattaj, I. W. *Nature* **345**, 502–506 (1990).
- Nagai, K., Oubridge, C., Jessen, T.-H., Li, J. & Evans, P. R. *Nature* **348**, 515–520 (1990).
- Jessen, T.-H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. *EMBO J.* **10**, 3447–3456 (1991).
- Scherley, D., Dathan, N. A., Boelens, W., van Venrooij, W. J. & Mattaj, I. W. *EMBO J.* **9**, 3675–3681 (1990).
- Lutz-Freyermuth, C., Query, C. C. & Keene, J. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6393–6397 (1990).
- Hall, K. B. *Biochemistry* **33**, 10076–10088 (1994).
- Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. *EMBO J.* **13**, 3873–3881 (1994).
- Hoffmann, D. W., Query, C. C., Golden, B. L., White, S. W. & Keene, J. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2495–2499 (1991).
- Habets, W. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2421–2425 (1987).
- Wittekind, M., Görlich, M., Friedrichs, M., Dreyfuss, G. & Mueller, L. *Biochemistry* **31**, 6254–6265 (1992).
- Görlich, M., Wittekind, M., Beckman, R. A., Mueller, L. & Dreyfuss, G. *EMBO J.* **11**, 3289–3295 (1992).
- Boelens, W. C. et al. *Cell* **72**, 881–892 (1993).
- van Gelder, C. W. G. et al. *EMBO J.* **12**, 5191–5200 (1993).
- Usman, N., Ogilvie, K. K., Tiang, M.-Y. & Cedergren, R. G. *J. Am. chem. Soc.* **109**, 7845–7854 (1987).
- Nagai, K., Evans, P. R., Li, J. & Oubridge, C. in *Isomorphous Replacement and Anomalous Scattering*, Proc. CCP4 St. W. (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 141–149 (Daresbury Laboratory, Warrington, UK, 1991).
- Tronrud, D. E., TenEyck, L. F. & Matthews, B. W. *Acta crystallogr.* **A43**, 489–501 (1987).
- Tsai, D. E., Harper, D. S. & Keene, J. D. *Nucleic Acids Res.* **19**, 4931–4936 (1992).

- Görlich, M., Burd, C. G. & Dreyfuss, G. *J. biol. Chem.* **269**, 23074–23078 (1994).
- Query, C. C., Bentley, R. C. & Keene, J. D. *Cell* **57**, 89–101 (1989).
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. *Nature* **365**, 512–519 (1993).
- Kim, J. L., Nikolov, D. B. & Burley, S. K. *Nature* **365**, 520–527 (1993).
- Burd, C. G. & Dreyfuss, G. *EMBO J.* **13**, 1197–1204 (1994).
- Görlich, M., Burd, C. G. & Dreyfuss, G. *Exp. Cell Res.* **211**, 400–407 (1994).
- Roud, M. A., Perona, J. J., Söll, D. & Steitz, T. A. *Science* **246**, 1135–1142 (1989).
- Ruff, M. et al. *Science* **252**, 1682–1689 (1991).
- Beirli, H. et al. *Science* **263**, 1432–1436 (1994).
- Roud, M. A., Perona, J. J. & Steitz, T. A. *Nature* **352**, 213–218 (1991).
- Cavarelli, J., Rees, B., Ruff, M., Thierry, J.-C. & Moras, D. *Nature* **362**, 181–184 (1993).
- Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. *Nature* **371**, 623–626 (1994).
- Lindahl, M. et al. *EMBO J.* **13**, 1249–1254 (1994).
- Sakabe, N. *Nucl. Inst. Meth.* **A303**, 448–463 (1991).
- Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* No. 26 (Daresbury Laboratory, Warrington, UK, 1992).
- CCP4 suite *Acta crystallogr.* **D50**, 760–763 (1994).
- Cosier, J. & Glazer, A. M. *J. appl. Cryst.* **19**, 105–107 (1986).
- Teng, T. *J. appl. Crystallogr.* **23**, 387–391 (1990).
- Sheldrick, G. M. in *Isomorphous Replacement and Anomalous Scattering*, Proc. CCP4 St. W. (eds Wolf, W., Evans, P. R., & Leslie, A. G. W.) 23–38 (Daresbury Laboratory, Warrington, UK, 1991).
- Otwiński, Z. in *Isomorphous Replacement and Anomalous Scattering*, Proc. CCP4 St. W. (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (Daresbury Laboratory, Warrington, UK, 1991).
- Abrahams, J. P., Leslie, A. G. A., Lutter, R. & Walker, J. E. *Nature* **370**, 621–628 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
- Krol, A. et al. *Nucleic Acids Res.* **18**, 3803–3811 (1990).
- Brünger, A. T. *Nature* **335**, 472–475 (1992).
- Gait, M. J., Pritchard, C. & Slim, G. in *Oligonucleotide Synthesis: A Practical Approach* (ed. Eckstein, F.) 25–48 (IRL, Oxford, 1991).

ACKNOWLEDGEMENTS. We thank the staff of the Synchrotron Radiation facility at Daresbury Laboratory and N. Sakabe, S.-Y. Park and A. Nakagawa at the Photon factory for help with data collection; J. Li, C. Pritchard, T. Smith, S. Iwai and J. Schwabe for help and advice; R. Dutton for illustrations; and B. Luisi, W. Sundquist and members of the LMB for criticism. This work was supported by the MRC and the Human Frontiers Science Program. The atomic coordinates will be deposited with the Brookhaven Protein Data Bank after further refinement.

An X-ray temperature map of the merging galaxy cluster A2256

Ulrich G. Briel* & J. Patrick Henry*†

* Max-Planck-Institut für extraterrestrische Physik, 85740 Garching, Germany

† Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

RICH clusters of galaxies are the largest bound masses in the Universe, but it has been difficult to determine just how massive they are^{1,2}. Observations of the velocities of individual galaxies, whose motions reflect the gravitational potential that they feel, have the difficulty that only the component of motion along the line of sight can be measured—the transverse component is unknown. X-ray observations seem to provide a better estimate of the mass^{3–6}, under the assumption that the hot X-ray-emitting gas is relaxed because of its isotropic velocity dispersion, but they have been hampered by a lack of adequate simultaneous spatial and spectral resolution. Here we present a map of the distribution of X-ray temperatures in the cluster Abell 2256, which has been thought to be nearly relaxed. Our map demonstrates that this is not the case, and implies that the assumption that the cluster is relaxed will lead to an underestimate of the true mass. As most clusters show still less evidence of being relaxed than A2256, their mass estimates may have larger errors.

A2256 was one of the first targets of the Rosat X-ray observatory during its calibration and verification phase, when it was observed for 17,323 s with the positional sensitive proportional counter (PSPC). These observations and their significance have been described in refs 5, 7 and 8. Four follow-up PSPC observations, together with the above-mentioned pointing, have yielded a total integration time on the cluster of 55,757 s. These new pointings were placed symmetrically around the cluster centre at an offset of ~ 20 arcmin, well within the 60-arcmin-radius field of view of the PSPC. Point sources in overlapping regions were used to align the individual pointings to an accuracy of ~ 2 arcsec; the necessary shifts were less than 10 arcsec in both coordinates, in accordance with the pointing accuracy of the Rosat spacecraft.

Figure 1 shows a contour plot of the X-ray surface-brightness distribution of the summed pointings in the energy band 0.5–2.5 keV. This contour plot confirms our previous finding of two maxima in the centre of the cluster. It shows very well the compressed contours south of the western maximum, which we interpreted as the contact plane of a small infalling group as its intragroup gas ‘snow-ploughs’ through the main intracluster gas⁷. The figure also shows the diffuse X-ray emission of the cluster extending further out than previously seen. To determine the total extent we binned the data in rings, excluded point sources as well as the region with the subgroup, and fitted these data to the β -model⁹ (see Fig. 2). This model assumes that the gas density is spherically symmetric and is given by $n_{\text{gas}}(r) = n_0[1 + (r/a_x)^2]^{-3\beta/2}$ where r is the distance from the cluster centre, n_0 is the central electron density and a_x is the physical core radius. This gives a projected angular surface brightness of $I(\theta) = I_0[1 + (\theta/\theta_x)^2]^{-3\beta+1/2}$ where θ is the angular distance from the cluster centre, I_0 is the central surface brightness, and θ_x is the angular radius corresponding to a_x . Under certain assumptions¹⁰, β is the ratio of galaxy to gas kinetic energies per unit mass. We obtained for the best fit $\theta_x = 5.35 \pm 0.11$ arcmin or $a_x = 0.271 \pm 0.006h^{-1}$ Mpc, $\beta = 0.81 \pm 0.01$ and $I_0 = 0.0268 \pm 0.0005$ counts s^{-1} arcmin $^{-2}$ (68% joint confidence level for three parameters), in good agreement with the values obtained previously⁵. We parametrize our ignorance of the scale of the Universe by setting the Hubble constant, the ratio of

recessional velocity to distance, to $H_0 = 100h$ km s^{-1} Mpc $^{-1}$. Recent determinations¹¹ yield $h \approx 0.9$. The diffuse emission could be traced out to a radius of 60 arcmin, corresponding to $3h^{-1}$ Mpc, which is by far the largest volume of intracluster gas ever observed directly in X-rays.

Measuring a true two-dimensional temperature map of a galaxy cluster has not been done before because previous imaging X-ray telescopes (on the Einstein observatory¹² or the Exosat observatory¹³) had insufficient spectral resolution, whereas non-imaging instruments had insufficient spatial resolution. The PSPC on the Rosat observatory^{14,15} is the first instrument with performance adequate to obtain such a map. Nevertheless, the procedure that we have used is rather complicated because of the five pointings and so we give a brief description of it. A more detailed description of parts of the procedure is given by Snowden *et al.*^{16,17}.

All the data making up the map came from observation times when the high energy and particle background rate was low.

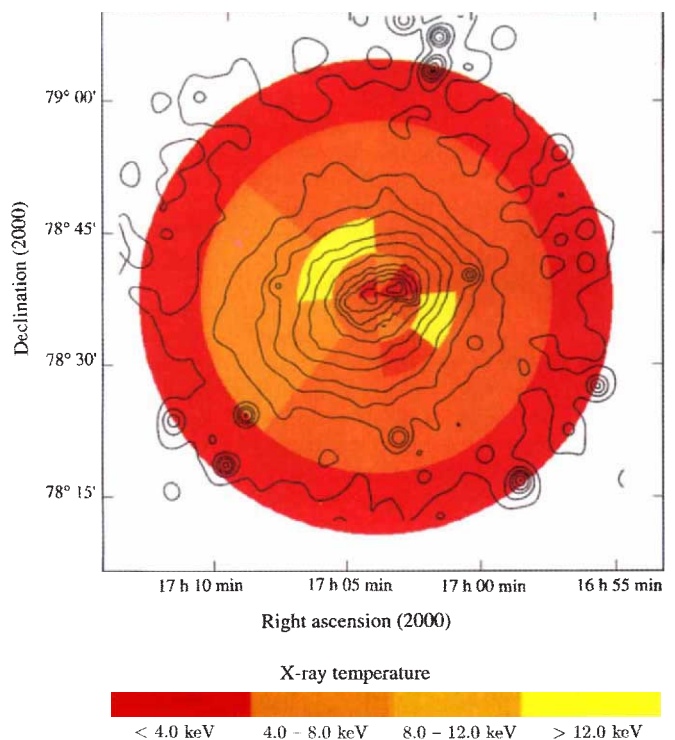


FIG. 1 Temperature map of the X-ray-emitting gas in the Abell cluster of galaxies A2256, deduced from pointed observations with the PSPC of the Rosat observatory. In this figure, North is up and East is to the left. The plasma temperatures are colour-coded in four temperature bands, which are separated by $\sim 1\sigma$ (for one parameter), except for the outermost ring ($3.9^{+0.0}_{-1.5}$ keV) and for the two high-temperature regions (hotter than 12 keV (northeast) and 9 keV (southwest) at 68% confidence). Overlaid are isointensity contours of the X-ray surface brightness of the summed pointings in the energy band 0.5–2.5 keV. To increase the overall spatial resolution in the image, only photons with an off-axis angle < 20 arcmin were used from the individual pointings. The image has been corrected for vignetting and for the differences in exposure between the various pointings and has been smoothed with a two-dimensional gaussian filter with $1\sigma = 15$ arcsec above a level of 5×10^{-4} counts s^{-1} pixel $^{-1}$ (1 pixel has a size of 15×15 arcsec 2); below that flux level, the image has been smoothed with gaussian filters of variable size to visualize better the low surface-brightness of the cluster. The contours are at the following levels: 0.2, 0.4, 0.7, 1.1, 2.0, 3.0, 5.0, 8.0, 11.0, 14.0, 17.0 and 20.0×10^{-4} counts s^{-1} pixel $^{-1}$. The background for this image, which was determined from point-source-free regions ~ 60 arcmin from the cluster centre, was 0.14×10^{-4} counts s^{-1} pixel $^{-1}$. The conversion of count-rate to flux, corrected for absorption by the Milky Way, is 1.55×10^{-11} erg cm^{-2} per PSPC-count.

For each pointing, the total background was determined from regions that were free of point sources and were 50–55 arcmin from the cluster centre. The contamination of the background by the cluster emission in those regions was $<15\%$. The detector gain during the five pointings was individually adjusted such that in each case the temperature of the cluster measured within 30 arcmin agreed with the Ginga measurement of 7.5 keV (ref. 18) when fitted with the Raymond–Smith plasma X-ray-emission model¹⁹ with heavy-element abundances fixed to the Ginga value¹⁸. The gain adjustments were all $<1\%$ of nominal.

We performed spectral fits to data from sectors of rings centred near the eastern peak. The precise regions were chosen to provide a balance between spatially resolving interesting features while still providing enough photons to measure the temperature accurately. None of the final regions chosen overlapped any other. The background-subtracted spectra for a given region from each pointing were summed, after correcting each photon for vignetting by an energy- and position-dependent factor (see ref. 20 for details) and then fitted to Raymond–Smith models with the heavy-element abundance fixed to 0.28 of their solar values. Free parameters were the temperature, absorption by material in the Milky Way along the line of sight to A2256 and normalization.

Figure 1 shows the results; there are three main points to note. (1) The infalling subgroup has a low temperature of $3.7^{+1.7}_{-1.0}$ keV (68% confidence for one parameter), confirming the results in ref. 7. (2) There are two hot regions, opposite each other and almost perpendicular to the presumed infall direction of the subgroup. These two regions are hotter than the cluster average at 90% (southwest) and 99% (northeast) confidence. (3) The rest of the intracluster gas is nearly isothermal with a slight indication of a falling temperature towards the outskirts of the cluster. This point is shown more clearly in Fig. 3, where we give the azimuthally averaged temperatures as a function of radius.

Although previous optical^{21,23} and X-ray²⁴ observations suggested the possibility of substructure, which is usually interpreted as being due to the merging of a group of galaxies and its associated gas with the main cluster, our data provide unequivocal evidence for a merger event. In most simulations of cluster

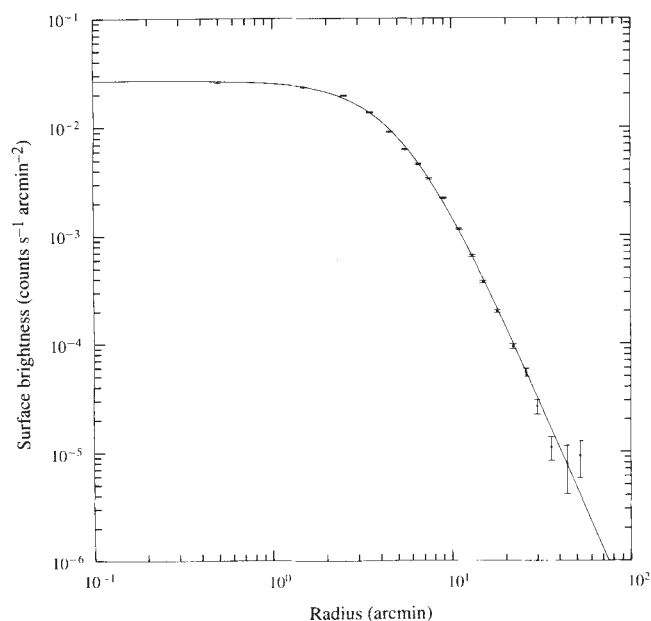


FIG. 2 Azimuthally averaged surface brightness of A2256 in the 0.5–2.5 keV band. The data points are plotted in the middle of each radial bin. The smooth curve is the best fitting β -model (see text). Error bars are at the 68% confidence level.

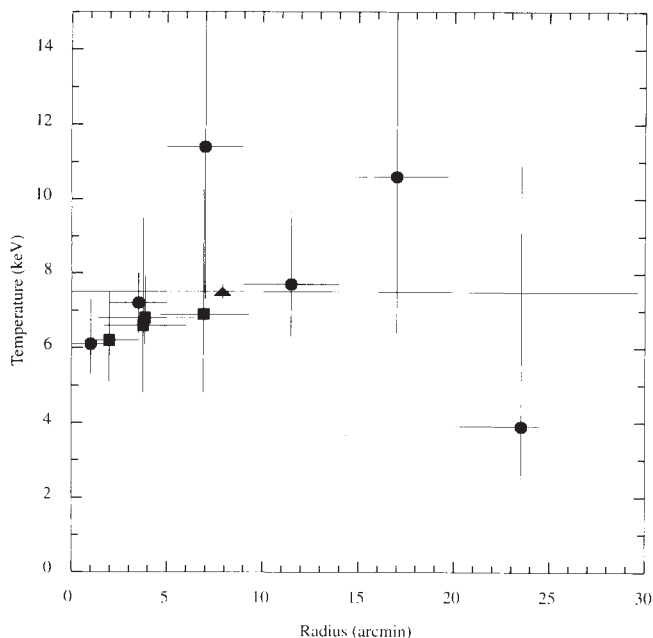


FIG. 3 Azimuthally averaged temperatures of A2256 from several instruments for the regions indicated. Filled circles, this work; filled squares, Broad Band X-Ray Telescope data³²; filled triangle, Ginga wide-beam measurement¹⁸.

formation^{25,27}, two hotspots occur at the time of the first penetration of the group into the main cluster when two jets of hot gas squirt out perpendicular to the merger axis. Because we do observe the hotspots approximately perpendicular to the independently determined merger axis (from the compressed surface-brightness contours), we conclude that the temperature map supports the merger hypothesis and provides an indication of how far the merger has progressed. The simulation that appears most similar to A2256 is given in ref. 26, where another ~ 1 Gyr will elapse before closest approach of the two components and complete relaxation of the gas occurs ~ 2 Gyr after that.

If the cluster is approximately isothermal at the Ginga temperature, as is indicated by our data, then its total mass within 27 arcmin ($=1.4h^{-1}$ Mpc), the largest radius with temperature information, is $\sim (9.5 \pm 0.3) \times 10^{14} h^{-1} M_{\odot}$. The error here is just the 68% confidence-level statistical error. This value agrees with our previous determination, but has a substantially smaller error because the temperatures have been measured out to a larger radius. The systematic error resulting from the unknown binding-mass distribution is about 20% of the measured value for radii where temperature information exists⁵. As we discuss below, the systematic error from the incomplete gas relaxation of this cluster is of the same order.

Using the formula in ref. 28, we find that the gas mass within the same radius is $(6.33 \pm 1.17) \times 10^{13} h^{-5/2} M_{\odot}$, a value which does not depend on the cluster being relaxed, only on the correctness of the β -model. Traditionally, astronomers also calculate the visible light from all the galaxies in a cluster. We have shown⁵ that this quantity is very uncertain in A2256, so we do not give it here. We note that the mass from the galaxies, although similarly uncertain, is small compared to that from the gas. The ratio of gas mass to total mass, which is a lower limit on the baryon mass to total mass ratio, is $(0.07 \pm 0.01)h^{-3/2}$. This large value is consistent with that found in the Coma cluster²⁹ and leads to a similar challenge to the standard cosmological picture as described in ref. 30 for Coma.

Our data also imply a reassessment of the long-held hope that spatially resolved cluster X-ray temperature measurements

would enable the measurement of cluster masses free of the systematic uncertainties of such measurements using optical methods. Both methods assume that the clusters are in equilibrium. Our data indicate that a nonequilibrium situation exists in this cluster, which leads to an underestimate of the cluster mass because of the neglected additional pressure from the bulk motion of the merger. Fortunately, A2256 does not seem to be far from equilibrium because the temperatures are approximately isothermal, so the derived total mass is probably accurate at the 30% level^{27,31}. However, A2256 is one of the most regular-looking clusters known. It is not at all reassuring that such an object is in fact in the process of a merger because there are many other clusters which are less regular in appearance. The systematic error on a mass estimate assuming relaxation will probably be large for these objects. What is needed now are realistic simulations of mergers that are adjusted to match both the X-ray surface brightness and spatially resolved temperature measurements of individual objects. The mass of the cluster, or at least an estimate of the systematic error, could possibly be extracted from these simulations on a case-by-case basis. A2256 would be an excellent place to start. □

Received 5 September; accepted 19 October 1994.

1. Merritt, D. *Astrophys. J.* **313**, 121–135 (1987).
2. The, L. & White, S. D. M. *Astr. J.* **92**, 1248–1253 (1986).
3. Fabricant, D. & Gorenstein, P. *Astrophys. J.* **267**, 535–546 (1983).

4. Hughes, J. *Astrophys. J.* **337**, 21–33 (1989).
5. Henry, J. P., Briel, U. G. & Nulsen, P. E. J. *Astr. Astrophys.* **271**, 413–420 (1993).
6. Allen, S. W. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **269**, 409–426 (1994).
7. Briel, U. G. et al. *Astr. Astrophys.* **246**, L10–L13 (1991).
8. Henry, J. P. & Briel, U. G. *Astr. Astrophys.* **246**, L14–L16 (1991).
9. Jones, C. & Forman, W. *Astrophys. J.* **276**, 38–55 (1984).
10. Bahcall, N. A. & Lubin, L. M. *Astrophys. J.* **426**, 513–515 (1994).
11. Pierce, M. J. et al. *Nature* **371**, 385–389 (1994).
12. Giacconi, R. et al. *Astrophys. J.* **230**, 540–550 (1979).
13. Taylor, B. G. et al. *Space Sci. Rev.* **30**, 479–512 (1981).
14. Trümper, J. *Adv. Space Res.* **2**, 241–249 (1983).
15. Pfeffermann, E. et al. *Proc. Soc. Photo-Opt. Instrum. Engrs* **733**, 519–532 (1986).
16. Snowden, S. L., McCammon, D., Burrows, D. N. & Mendenhall, J. A. *Astrophys. J.* **424**, 714–728 (1994).
17. Snowden, S. L., Plucinsky, P. P., Briel, U., Hasinger, G. & Pfeffermann, E. *Astrophys. J.* **393**, 819–828 (1992).
18. Hatsukade, I. thesis, Osaka Univ. (1989).
19. Raymond, J. C. & Smith, B. W. *Astrophys. J. Suppl. Ser.* **35**, 419–439 (1977).
20. Zimmermann, H. U., Belloni, T., Izzo, C., Kahabka, P. & Schwendker, O. *Max-Planck-Institut für extraterrestrische Physik Rep.* 244 (Garching, 1993).
21. Geller, M. J. *Comments Astrophys.* **10**, 47–52 (1984).
22. Geller, M. J. & Beers, T. C. *Publ. astr. Soc. Pacif.* **94**, 421–439 (1982).
23. Oegerle, W. R., Hoessel, J. G. & Jewison, M. S. *Astr. J.* **93**, 519–528 (1987).
24. Fabricant, D., Kent, S. & Kurtz, M. *Astrophys. J.* **336**, 77–92 (1989).
25. Schindler, S. & Müller, E. *Astr. Astrophys.* **272**, 137–152 (1993).
26. Evrard, G. in *Clusters of Galaxies* (eds Oegerle, W. R., Fitchett, M. J. & Daney, L.) 287–326 (Cambridge Univ. Press, 1990).
27. Evrard, G. in *Clusters of Galaxies* (ed. Durret, F.) (Frontières, Gif-sur-Yvette, in the press).
28. Henry, J. P. & Henriksen, M. J. *Astrophys. J.* **301**, 689–697 (1986).
29. Briel, U. G., Henry, J. P. & Böhringer, H. *Astr. Astrophys.* **259**, L31–L34 (1992).
30. White, S. D. M., Navarro, J. F., Evrard, A. E. & Frenk, C. S. *Nature* **366**, 429–433 (1993).
31. Navarro, J. F., Frenk, C. S. & White, S. D. M. *Mon. Not. R. astr. Soc.* (submitted).
32. Miyaji, T. et al. *Astrophys. J.* **419**, 66–77 (1993).

ACKNOWLEDGEMENTS. J.P.H. thanks J. Trümper and the Rosat group for their hospitality during the course of this research. The Rosat project is supported by the Bundesministerium für Forschung und Technologie. J.P.H. was supported by NASA, NATO and the US NSF.

Secular resonances and cometary orbits in the β Pictoris system

Harold F. Levison*, Martin J. Duncan† & George W. Wetherill‡

* Space Science Department, Southwest Research Institute, 6220 Culebra Road, San Antonio, Texas 78238, USA

† Physics Department, Queen's University, Kingston, Ontario K7L 3N6, Canada

‡ Carnegie Institution of Washington, Department of Terrestrial Magnetism, Washington DC 20015, USA

THE dusty disk around the star β Pictoris is believed to contain a large number of comet-like bodies^{1,2}. Transient absorption events associated with β Pictoris have been observed at many wavelengths^{3–6} and attributed to cometary objects on eccentric orbits passing between the star and the Earth⁷. One unexplained aspect of these events is the large asymmetry between red-shifted and blue-shifted features: ~90% of the events are red-shifted⁶. We show here that such an asymmetry is a natural consequence of the influence on cometary orbits of secular resonances associated with some planetary systems. Results from numerical integrations of test particles in a model of our Solar System show that the ν_6 secular resonance excites objects initially on near-circular orbits to high eccentricities while aligning their perihelia. Depending on the location of a distant observer, this alignment can produce an asymmetry similar to that observed for β Pictoris. Our results imply the presence of at least two planets (a prerequisite for the existence of secular resonances) around β Pictoris.

Data from the Infrared Astronomy Satellite (IRAS) indicate that most A, F and G stars possess dusty disks⁸. However, β Pictoris is the only known star whose disk has been imaged directly⁹. This disk extends at least 1,000 AU from the star and may have an inner cleared zone similar to the size of our planetary system (~3–30 AU). The disk is seen nearly edge-on ($\leq 14^\circ$)

and is very flat (its opening angle is $\sim 10^\circ$)¹⁰. Because β Pictoris is much older than the dynamical lifetime of the dust surrounding it, larger objects must be creating that dust today. It has been suggested¹ that colliding comets are the source of this dust.

One of the interesting aspects of the β Pictoris system is the existence of Doppler-shifted, transient absorption features observed in many spectral lines^{3–5}. These absorption events have velocities relative to the star that can be as high as $\sim 300 \text{ km s}^{-1}$ (comparable to the escape velocity near the star's surface), but typically are $\sim 20\text{--}30 \text{ km s}^{-1}$ (comparable to the orbital velocity of an object near 1 AU). One of the most puzzling aspects of these events is the large asymmetry between red-shifted and blue-shifted features: the fraction of events that are red-shifted is ~ 0.9 (ref. 6).

The transient absorption events have been interpreted as being due to cometary objects on eccentric orbits passing between the star and the Earth⁷. Such objects would produce the observed absorption lines when they are within ~ 20 stellar radii of the star. This interpretation does not readily explain the Doppler-shift asymmetry, however. If the events are due to comets, then some mechanism must be aligning the orbits to produce the asymmetry. The most highly developed model to date involves the scattering of many comets, resulting from the breakup of a single Chiron-sized ($\sim 100 \text{ km}$ radius) object, by a planet on a very eccentric ($e \geq 0.6$) orbit^{2,11}. This model seems improbable.

Here we present the results from numerical integrations of massless test particles in our Solar System. These results show that it is possible to reproduce, as a result of secular resonances, many of the qualitative features of β Pictoris' transient absorption events. We do not attempt to reproduce the absorption events in detail because β Pictoris' planetary system is undoubtedly different from our own. Our approach is essentially a plausibility argument to demonstrate that a solar system with two or more planets can excite some objects from circular to highly eccentric orbits while aligning their perihelia in such a way to produce a Doppler-shift asymmetry.

Secular effects in planetary systems arise physically from the fact that the first-order effect of one body on another (averaged

over the orbital periods) causes the orbital precession both of the longitude of perihelion, $\tilde{\omega}$ (approximately the direction to the point at which the body is closest to the star), and of the longitude of the ascending node, Ω (the direction to the point where the body crosses the plane of the solar system in an 'upward' direction). For a given planet the precession of $\tilde{\omega}$ is typically dominated by one frequency¹². The same is true for Ω , although the dominant frequency is usually different. Thus there are $2N$ fundamental frequencies in a planetary system with N planets. The periods associated with these frequencies range from 46,000 to 2×10^6 years in our Solar System—much larger than the orbital periods of the planets.

The orbit of a test particle placed in the above planetary system will also precess in $\tilde{\omega}$ and Ω because of the effects of the planets. If one of the precession frequencies of the test particle is commensurate with one of fundamental frequencies of the planetary system, a 'secular resonance' is said to occur. Secular resonances have played a crucial role in determining the structure of our Solar System (see ref. 13 for a review). Note that a secular resonance requires the precession of planetary orbits, which means that there must be at least two planets in the system.

The particular resonance used in our plausibility argument is the ν_6 resonance, which is located at 2.1 AU from the Sun in our Solar System and has a frequency of $g_6 = 2\pi/46,000 \text{ yr}^{-1}$. Williams¹⁴, Yoshikawa¹⁵ and Morbidelli¹⁶ have shown that the

eccentricity of a typical test particle very near this secular resonance oscillates up and down and that the critical angle for this resonance, $\sigma_6 \equiv \tilde{\omega} - g_6 t$ where t is time, librates about either 0° or 180° . Thus these resonances increase the eccentricity of the orbits of the particles that librate around them and align them with the orientation of Saturn's orbit. (As Saturn's longitude of perihelion, $\tilde{\omega}_S$, librates about $g_6 t$, we define $\tilde{\omega}_S \equiv g_6 t$ as Saturn's mean longitude of perihelion.)

To test whether secular resonances could indeed produce the transient events observed at β Pictoris, we numerically integrated the orbits of 1,000 test particles in our Solar System. The particles were in initially low-eccentricity, low-inclination orbits near the ν_6 secular resonance at 2.1 AU. The orbits of the Sun and the four jovian planets were integrated as a full N -body system. The test particles themselves were not gravitationally interacting with each other. The equations of motion for the Sun, planets and particles were integrated in three dimensions using the powerful technique introduced by Wisdom and Holman¹⁷.

The test particles were initially placed on orbits lying in the invariable plane of the Solar System, with eccentricity $e=0.03$. Their initial longitude of perihelia and mean anomalies were randomly selected between 0 and 2π . Their initial semimajor axes, a , were uniformly distributed between 2.05 and 2.15 AU. The orbit of each test particle was followed until the particle (1) suffered a close encounter with a planet, (2) became a Sun-grazer (that is, passed within 2 solar radii of the Sun), or (3) survived for the length of the integration, $5 \times 10^6 \text{ yr}$. Particles on either side of this band in semimajor axis did not exhibit the large variations in eccentricity found near the resonance and are not discussed here.

At the end of the integration, 47% of the particles remained, 49% suffered a close encounter with Jupiter, and 4% became Sun-grazers. This fraction of Sun-grazers is not unusual as it has been found that $\sim 6\%$ of short-period comets¹⁸ and $\sim 50\%$ of near-Earth asteroids¹⁹ will end their lives by becoming Sun-grazers. Viewed from outside the Solar System, the velocity of

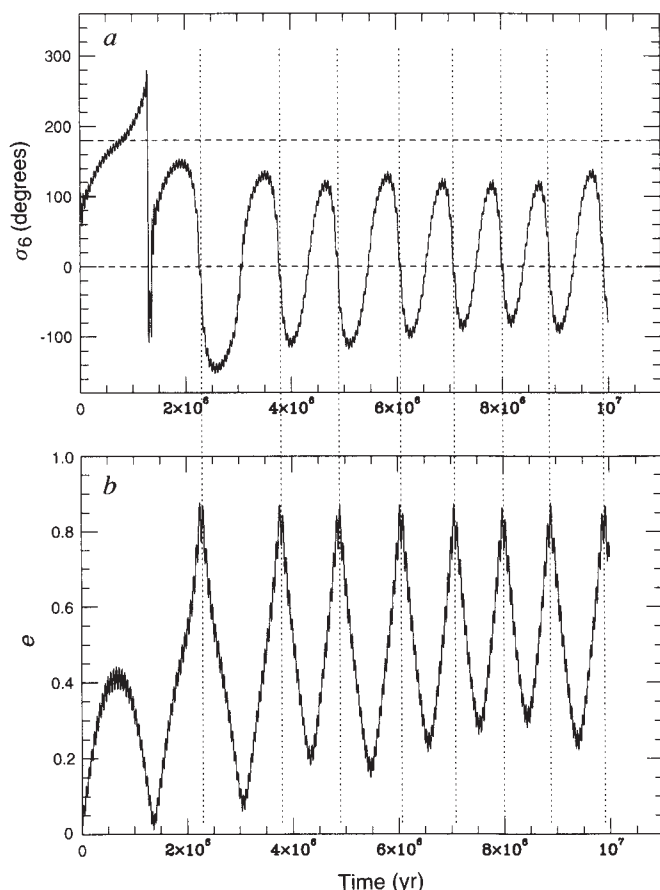


FIG. 1 The temporal evolution is shown of a typical particle displaying high-eccentricity spikes (capable of producing the absorption events if it were in the β Pictoris system). The particle is in the relevant mode after $\sim 4 \times 10^5 \text{ yr}$. While in this mode, the particle's eccentricity varies from ~ 0.2 to ~ 0.9 . a, Critical angle σ_6 librates about 0° with an amplitude of $\sim 120^\circ$. b, Eccentricity plotted against time. Note that when the orbit is at its largest eccentricity, σ_6 is always 0° .

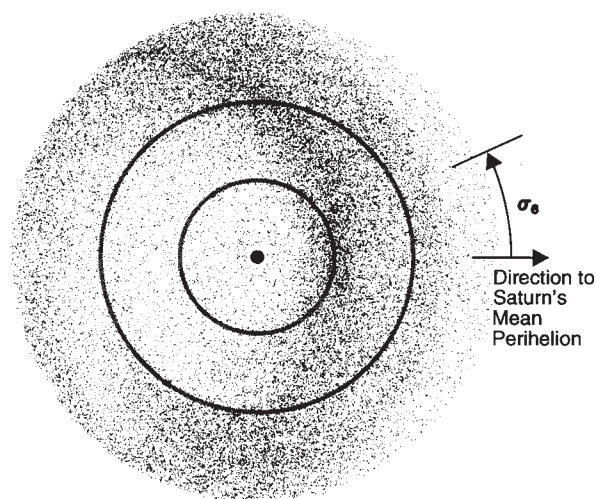


FIG. 2 The dots indicate the location every 10,000 years of the perihelion of each test particle in the integration with respect to the location of Saturn's mean perihelion. Only particles with $e > 0.7$ are plotted. The definition of the critical angle σ_6 is indicated. The inner and outer circles have radii of 0.2 AU and 0.4 AU respectively. There is a clear tendency for particles in the ν_6 secular resonances to fall along a curve that is closest to the Sun at $\sigma_6 = 0$. Thus the secular resonance aligns the orbits of particles. The particles that do not fall along this curve have been knocked out of the resonances by close approaches with Jupiter. If the secular resonance were closer to the Sun so that particles on nearly radial orbits could not encounter Jupiter (as may be the case at β Pictoris), then this alignment would be much stronger.

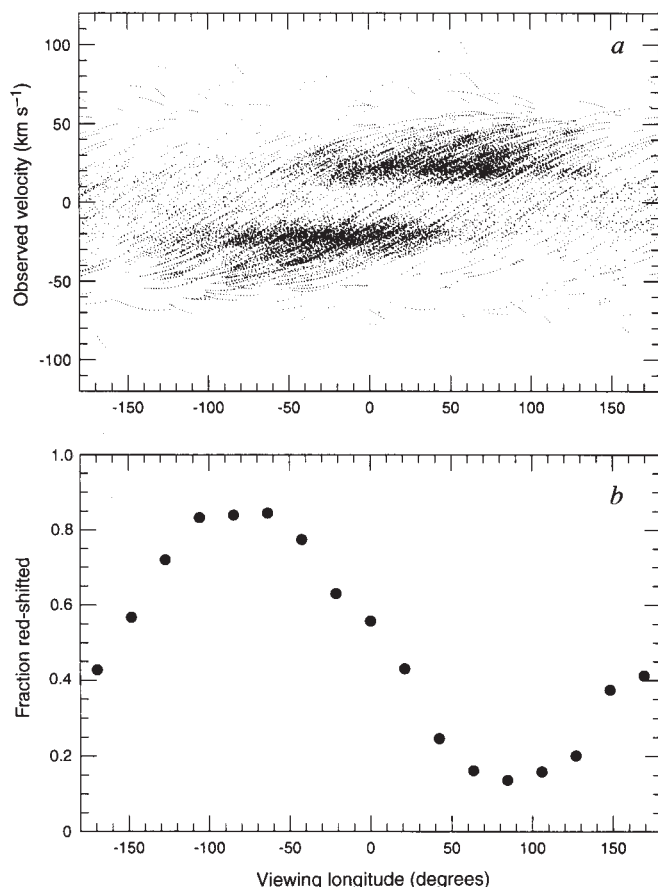


FIG. 3 *a*, Plots of what would be the observed radial velocity of the test particles (as projected in front of the star) as a function of the 'viewing longitude'. The 'viewing longitude' is defined as the angle between the position of the observer and the mean longitude of perihelion of the controlling planet of the secular resonance as seen from the centre of β Pictoris. Here the disk is assumed to be 5° from edge-on. In addition, in order to model the activity of comet-like bodies, we only allowed them to cause absorption lines when their distance from the star was <20 times the radius of β Pictoris. The results described in the text are not a strong function of the choice of this 'activation' distance. *b*, The fraction of red-shifted events as a function of viewing longitude. At the peak, at least 90% of the events are red-shifted.

the objects as they plunged into the Sun would be approximately equal to the escape velocity of the Sun, which is several hundred kilometres per second. The events would all be red-shifted because the objects would not survive the encounter. Sun-grazers produced in this way could explain the very high red-shifted transient events seen on β Pictoris.

We now concentrate on explaining the events with velocities ~ 20 – 50 km s $^{-1}$. The high-eccentricity objects that survived in our integration typically followed the two types of behaviour explained qualitatively in the semi-analytical work of Yoshikawa¹⁵. Of these, one produces objects with large enough eccentricities that they pass close enough to the Sun to produce atomic absorption events, which in the β Pictoris system is at 20 stellar radii. Figure 1 shows the temporal evolution of a typical particle showing this type of behaviour. The particle experiences periodic high-eccentricity ($e \approx 0.9$) spikes when $\sigma_6 = 0^\circ$. Figure 2 shows that when the orbits are at their peak eccentricity (where they would produce the absorption events if they were in the β Pictoris system), they are aligned.

To determine what effect the alignment of the perihelion of our test particle would have on the velocity distribution of the transient events, we calculate the absorption events that objects

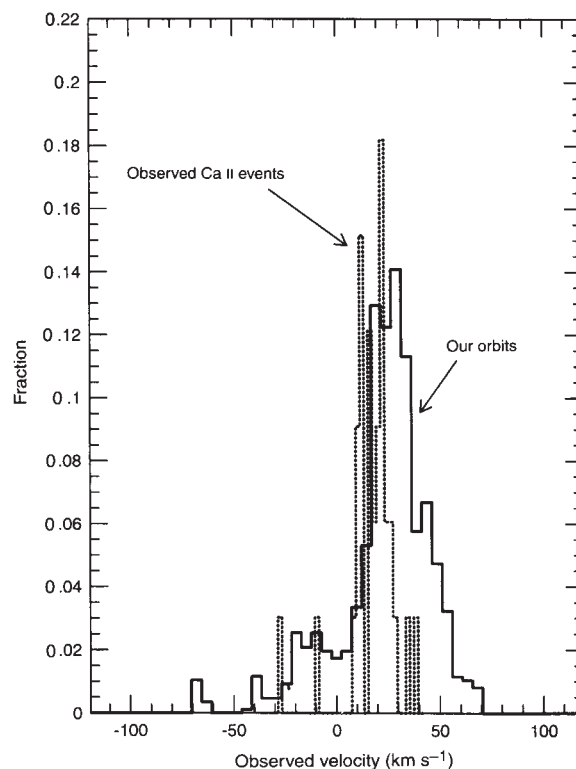


FIG. 4 Comparison of the histograms of the distribution of simulated and observed transient absorption event velocities. Dotted lines, the distribution of real Ca II events observed by Lagrange-Henri *et al.*⁶. Solid lines, the results of our calculation observed from a longitude of $\sim 90^\circ$ and a latitude of 5° . The two distributions are very similar, although our events appear to have slightly larger velocities. See text for further discussion.

in our calculations would produce if they were in the β Pictoris system. Recall that we are presenting the results from calculations performed in our Solar System, with the mass of the central star being one solar mass. Figure 3*a* shows the results of such a calculation in terms of the observed radial velocity of the test particle projected in front of the star as a function of the 'viewing longitude'. There is a significant asymmetry between the number of red-shifted and blue-shifted events for over half of the values of viewing longitude. This result is emphasized in Fig. 3*b*, which shows the fraction of red-shifted events as a function of viewing longitude. At the peak we should expect at least 90% of the events to be red-shifted. The observed asymmetry for events in the Ca II (singly ionized calcium) absorption line is also 90% (ref. 6).

A further comparison between our results and the observed events is presented in Fig. 4, which shows the distribution of event velocities. The two distributions are very similar, although our events appear to have slightly larger velocities. The higher velocity events may occur at smallest distances from the star so that all the Ca II has been ionized, or the secular resonance responsible for the events may be farther from β Pictoris than the v_6 resonance is from the Sun. In the latter case, the orbital velocities of the bodies would be smaller. With these caveats, the model presented here reproduces the observed absorption events very well with a planetary system similar to our own. $\square$

Received 4 August; accepted 19 October 1994.

1. Weissman, P. R. *Science* **224**, 987–989 (1984).
2. Beust, H., Vidal-Madjar, A., Ferlet, R. & Lagrange-Henri, A. M. *Astr. Astrophys.* **236**, 202–216 (1990).
3. Kondo, Y. & Bruhweiler, F. C. *Astrophys. J.* **291**, L1–L5 (1985).

4. Lagrange, A. M., Ferlet, R. & Vidal-Madjar, A. *Astr. Astrophys.* **173**, 289–292 (1987).
5. Ferlet, R., Vidal-Madjar, A. & Hobbs, L. M. *Astr. Astrophys.* **185**, 267–270 (1987).
6. Lagrange-Henri, A. M., Gosset, E., Beust, H. & Vidal-Madjar, A. *Astr. Astrophys.* **264**, 637–653 (1992).
7. Lagrange-Henri, A. M., Vidal-Madjar, A. & Ferlet, R. *Astr. Astrophys.* **190**, 275–282 (1988).
8. Aumann, H. H. *Astr. J.* **96**, 1415–1419 (1988).
9. Smith, B. A. & Terrie, R. J. *Science* **226**, 1421–1424 (1984).
10. Artymowicz, P., Burrows, C. & Paresce, F. *Astrophys. J.* **337**, 494–513 (1989).
11. Beust, H. & Lissauer, J. *Astr. Astrophys.* **282**, 804–810 (1994).
12. Brouwer, D. & Clemence, G. M. *Celestial Mechanics* Ch 16 (Academic, New York, 1961).
13. Duncan, M. & Quinn, T. A. *Rev. Astr. Astrophys.* **31**, 265–295 (1993).
14. Williams, J. G. thesis, Univ. California, Los Angeles (1969).
15. Yoshikawa, M. *Celest. Mech.* **40**, 233–272 (1987).
16. Morbidelli, A. *Icarus* **105**, 48–62 (1993).
17. Wisdom, J. & Holman, M. *Astr. J.* **102**, 1528–1538 (1991).
18. Levison, H. & Duncan, M. *Icarus* **108**, 18–36 (1994).
19. Farinella, P. et al. *Nature* **371**, 314–317 (1994).

ACKNOWLEDGEMENTS. We thank D. Backman and J. Lissauer for discussions. H.F.L. was supported by NASA, and M.J.D. was supported by the Canadian NSERC.

Light-emitting diodes with variable colours from polymer blends

M. Berggren*, O. Inganäs*, G. Gustafsson*, J. Rasmussen*, M. R. Andersson†‡, T. Hjertberg† & O. Wennerström†

* Laboratory of Applied Physics, University of Linköping, S-581 83 Linköping, Sweden

† Department of Organic Chemistry, ‡ Department of Polymer Technology, Chalmers University of Technology, S-412 96 Göteborg, Sweden

THE range of materials now available for polymer-based light-emitting diodes (LEDs) is such that electroluminescence can be obtained throughout the visible spectrum^{1–12}. Here we show that, by blending polymers with different emission and charge-transport characteristics, LEDs can be fabricated in which the emission colour varies as a function of the operating voltage. This phenomenon arises from the self-organizing properties of the blends, in which entropy drives phase separation of the constituent polymers and gives rise to submicrometre-sized domains having a range of compositions and emission characteristics. Emission from domains of different composition is controlled by the ease with which charge is injected, which in turn depends on the applied voltage.

The polymer light-emitting diodes used in this study are based on polymers with thiophene in the main chain. Polythiophene has been used in polymer LEDs^{13–16}: it is a polymer system with great chemical variability (through substitution at the 3- and 4-positions), which can easily be produced by 2,5-polymerization of the substituted monomer via oxidative polymerization. The purification of the polymers requires careful work, but is crucial for high-performance materials.

The design of our polymers is aimed at geometrical control of the polymer main chain. We can permanently control the torsion of the main chain, and thus the conjugation length, by altering the substitution pattern and character of substituents on the polythiophene chain. Enhanced steric hindrance from side groups leads to decreasing planarity of the main chain, and an increasing band gap; simple model calculations demonstrate how the band gap is enhanced on torsion¹⁷. Using this approach we have obtained electroluminescent polymer diodes based on different materials (I–IV in Fig. 1) with colours ranging from blue to near-infrared, with green, orange and red as intermediate steps (Fig. 1). In some of these we have incorporated so-called electron injection layers to enhance the efficiency^{18–20}. External quantum efficiencies are in the range 0.1–1%.

We find that LEDs incorporating blends of these polymers allow simultaneous emission from the different polymers in the blend. In blends of IV and II we observe a blue and a red emission peak; in blends of III and I we observe green and red emission peaks. The intensity ratio of these peaks is determined by the voltage applied to the diode, and the stoichiometry of the polymer blend. This therefore allows us to obtain a voltage-controlled variable-colour light source (Fig. 2). Similar voltage-controlled colours have recently been reported to occur at 80 K in a polymer/molecule blend LED²¹.

To understand this phenomenon, we have characterized the materials formed by spin-casting polymer blends, and studied polymer LEDs built from different blends. From optical absorption spectroscopy it is clear that the blends give optical absorption features from both polymers in the binary blends. Photoluminescence studies of the blends give clear evidence for partial exciton transfer from III to I, that is from the high-band-gap to the low-band-gap material. The efficiency of exciton transfer varies with the mole ratio of III to I, as described in the simple model of exciton diffusion by trapping of excitons on a guest in a host medium²¹. This exciton transfer (Förster resonance transfer) decays as R^{-6} , where R is the distance between dipoles. The transfer depends strongly on the relative orientation of dipoles, as well as on the overlap of emission and absorption spectra. The spectral overlap conditions are fulfilled for many combinations of these polymers.

No microstructure was observed by photomicroscopy or luminescence microscopy of the 1:1 mole ratio blends of II and III, I and III, and IV and II. We conclude that phase separation must lead to structures smaller than 1 μm , the limit to resolution of the optical microscope. Imaging spin-cast films of the blends using scanning force microscopy in both tapping and contact modes give clear evidence for phase separation (Fig. 3). As the fractions of the two different phases vary with the stoichiometry

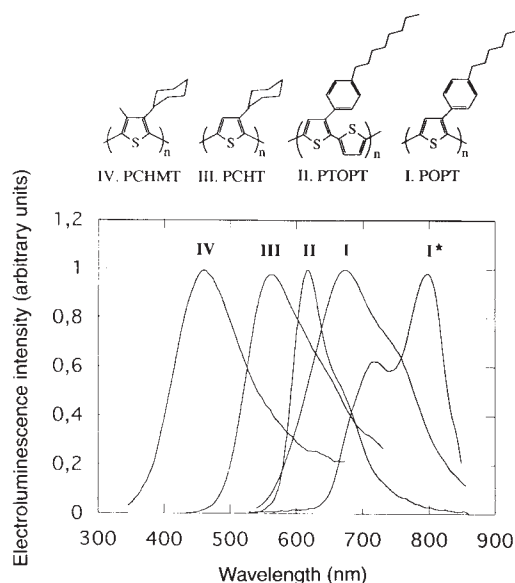


FIG. 1 Chemical structure and electroluminescence spectra of polymers I–IV. All of the diodes used for obtaining electroluminescence spectra were built using indium tin oxide (ITO) as the hole-injecting electrode and Ca/Al as the electron-injecting contact; luminescence spectra were obtained with an Oriel Instaspec 1B diode-array spectrometer. Polymer I exists in two forms and may be thermally converted from form I to form I* (ref. 20). The polymer diode of IV also contains a layer of 2-(4-biphenyl)-5-(4-tertbutylphenyl)-1,3,4-oxadiazole (PBD), which does not contribute to the electroluminescence.

of the polymer blend, we deduce that the smaller dark islands shown in the image contain the minority phase. We observe that at least one of the phases must be a pure homopolymer, as exciton transfer between the polymers is not complete. The observed dimension of the minority phase is 50–200 nm, depending on blend and stoichiometry. If interpreted as due to topography, the thickness variation of the polymer layer is at most 5 nm between the two phases. Scanning electron microscopy of blends formed by solvent casting also shows clear evidence for phase separation in the form of small spheres. Here the slow evaporation of solvent is expected to allow the polymer to reach phase structures closer to the thermodynamic equilibrium. It is therefore clear that phase separation occurs in the binary blends investigated so far. (Phase separation in polymer blends is also supported by consideration of the free-energy of mixing²².)

The thickness of the polymer layer in the polymer LED is typically 30–100 nm, which is of the same order as the width of the minority phase imaged by scanning force microscopy. We can thus imagine two different situations in the polymer-blend LEDs where phase separation occurs; the width of each domain is larger, or smaller, than the thickness of the layer. In the former case we find homopolymer phases extending from anode to cathode, essentially forming a great number of parallel-connected polymer LEDs. The observation that the fraction of the minority phase (as imaged by scanning force microscopy) roughly follows the stoichiometry of the polymer blend lends support to this interpretation. Electrical data for the polymer-blend LEDs do not clearly establish the presence of parallel connections of small individual polymer LEDs. It is conceivable that the statistics of many parallel diodes would completely suppress clear signs of two parallel processes in the current–voltage characteristics. More complex features may arise if homopolymer phases do not directly extend from anode to cathode. As long as the volume fraction of either polymer is not too low, however, we expect to be able to find a continuous but tortuous path of charge transport through one and the same type of polymer via bicontinuous percolation.

If we assume the model of parallel diodes with different voltage–luminance characteristics, we can easily understand how the voltage control of colour comes about. With fixed electrodes, barriers to the injection of both electrons and holes increase when the band gap increases²³ (as long as the workfunctions of the injecting electrodes are found inside the gap of the smallest-band-gap material). The turn-on voltage for electroluminescence increases in the order of polymers I to IV. The higher-energy

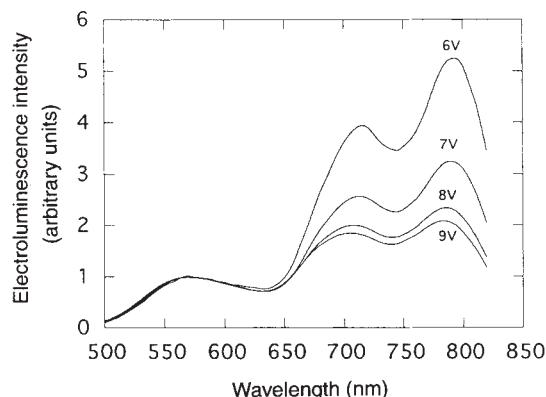


FIG. 2 Voltage control of electroluminescence in a I/III (1:2) blend. ITO and Ca/Al were used as hole- and electron-injecting contacts, respectively.

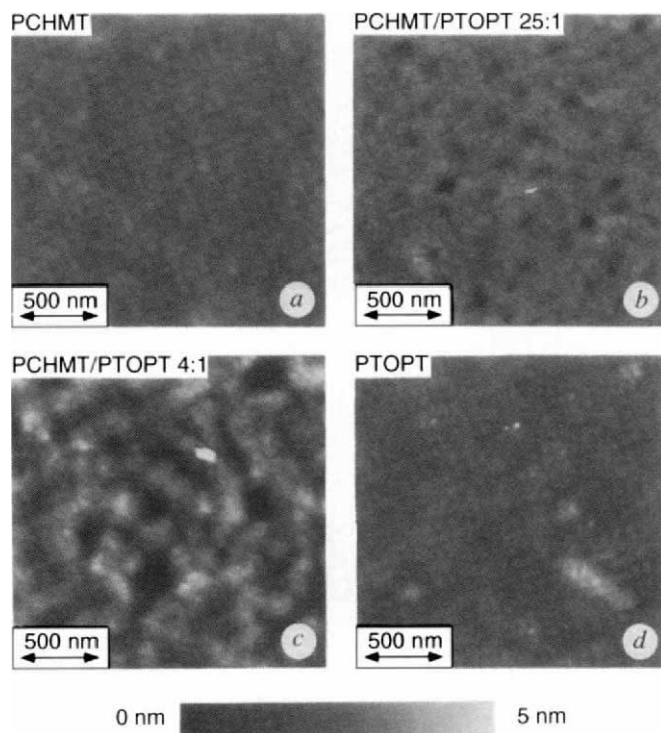


FIG. 3 Tapping-mode scanning force microscopy images of spin-cast films of pure IV (a), blend of 25:1 IV/II (b), a blend of 4:1 IV/II (c), and pure II (d). The contrast variation in all four images is indicated by the greyscale and covers 0–5 nm. The root mean square (r.m.s.) roughness of the images are 0.23, 0.412, 0.9 and 0.32 nm, that is, the polymer blends are rougher than the pure polymers. The images of the polymer blends also indicate that phase separation occurs; the observed dimension of the minority phase is 50–200 nm depending on blend and stoichiometry. Imaging by contact-mode scanning force microscopy gave very similar images and r.m.s. values.

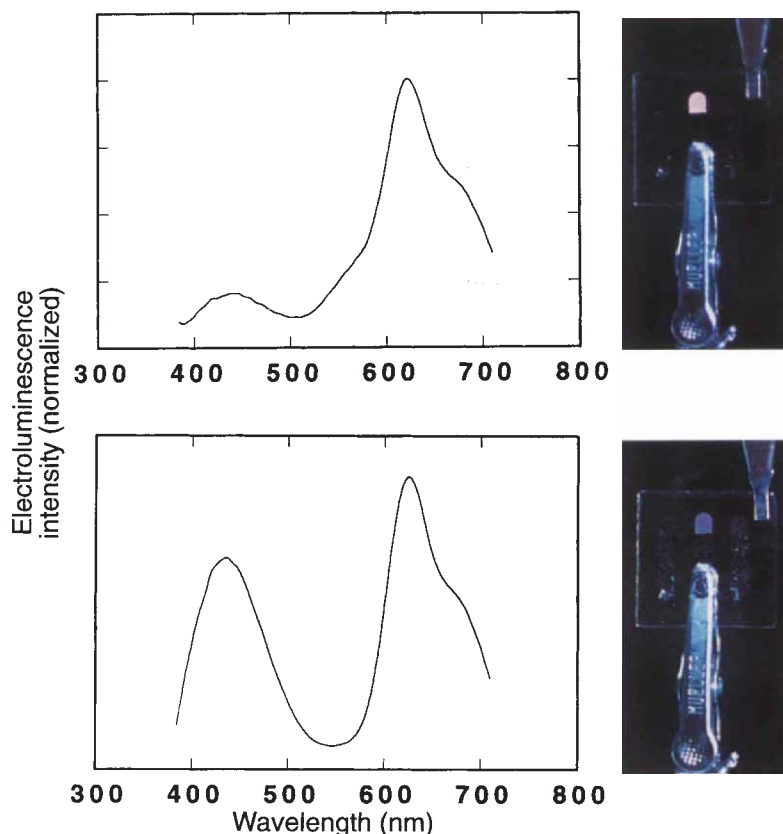
gap requires a higher field for electron–hole injection, and thus we expect the blue-shifted colours to start appearing at higher voltages, as observed. The thickness variation of 5–10% observed in scanning force microscopy is small enough to ignore the difference in the applied electric field in the submicrometre LEDs, at fixed voltage.

The limit to colour control is given by the presence of exciton transfer in the blends. This will always convert an amount of excitons from the higher-band-gap polymer to the one of lower band gap; we therefore have to add an excess of the higher-band-gap polymer. The fractions of excited states formed in this polymer, and lost by exciton transfer to the lower-band-gap polymer, will be highly dependent on the microstructure of the polymer blend. It is therefore attractive to control the size and shape of phases separating from these blends.

The technical possibilities of these polymer LEDs are demonstrated in Fig. 4. We can produce light sources of variable colour and intensity, as we can separately control the amplitude and duration of the voltage applied. We can easily mix colours such as red and blue, green and red, orange and blue, and expect also to be able to combine red, green and blue. This might possibly in the future lead to a polymeric electroluminescent colour screen. The simplicity of forming such a multicolour screen with passive addressing of individual multicolour pixels should be attractive.

We conclude that self-organization in polymer blends allows us to obtain voltage-controlled multiple-colour emission from polymer LEDs, an example of the new possibilities available with polymer electronics. Although the present generation of

FIG. 4 Voltage dependence of electroluminescence from a 50:1 IV/II blend in a PBD polymer-blend LED using a Ca/Al electrode on a polymer layer deposited by spinning on the ITO substrate. Top, at 22 V (100% duty cycle); bottom, at 28 V (20% duty cycle). At each voltage, the electroluminescence spectrum is shown, together with a photograph of the operating LED.



blend-polymer LEDs may not give the colour tunability necessary for a colour screen, we foresee the possibility of controlling phase separation in order to control or suppress exciton transfer, and we also expect to be able to develop new materials with reduced emission bandwidth, if necessary. The basic limits on the development of polymeric electroluminescent colour screens are more to be found in the instability of the devices, leading to degradation of polymer LEDs during operation. The degradation experienced by our multicolour devices is not dissimilar to that found to occur in the single-colour devices under development, and will benefit from progress in the latter field. □

Received 4 July; accepted 24 October 1994.

- Burroughes, J. H. *et al.* *Nature* **347**, 539–541 (1990).
- Brown, A. R. *et al.* *Appl. Phys. Lett.* **61**, 2793–2795 (1992).
- Burn, P. L. *et al.* *J. chem. Soc., chem. Commun.* 32–34 (1992).
- Braun, D. & Heeger, A. J. *Appl. Phys. Lett.* **58**, 1982–1984 (1991).
- Greenham, N. C., Moratti, S. C., Bradley, D. D. C., Friend, R. H. & Holmes, A. B. *Nature* **365**, 628–630 (1993).
- Kraft, A. *et al.* *Synth. Met.* **55**, 936–941 (1993).
- Zhang, C., Hoger, S., Pakbaz, K., Wudl, F. & Heeger, A. J. *J. electron. Mater.* **22**, 413–417 (1993).
- Vestweber, H. *et al.* *Adv. Mater.* **4**, 661–662 (1992).
- Grem, G., Leditzky, G., Ullrich, B. & Leising, G. *Adv. Mater.* **4**, 36–37 (1992).
- Grem, G. & Leising, G. *Synth. Met.* **57**, 4105–4110 (1993).
- Gustafsson, G. *et al.* *Nature* **357**, 477–479 (1992).
- Dyrekleiv, P. *et al.* *Adv. Mater.* (in the press).
- Ohmori, Y., Uchida, M., Muro, K. & Yoshino, K. *Solid. St. Commun.* **80**, 605–608 (1991).
- Braun, D., Gustafsson, G., McBranch, D. & Heeger, A. J. *J. appl. Phys.* **72**, 564–568 (1992).
- Greenham, N. C., Brown, A. R., Bradley, D. & Friend, R. H. *Synth. Met.* **57**, 4134–4138 (1993).
- Gill, R. E., Malliaras, G. G., Wildeman, J. & Hadzioannou, G. *Adv. Mater.* **6**, 132–135 (1994).
- Themans, B., Salaneck, W. R. & Bredas, J. L. *Synth. Met.* **28**, C359–C364 (1989).
- Berggren, M. *et al.* *Adv. Mater.* **6**, 488–490 (1994).
- Berggren, M. *et al.* *J. appl. Phys.* (1994).
- Berggren, M. *et al.* *Appl. Phys. Lett.* **65**, 1489–1491 (1994).
- Agranovich, V. M. & Galanin, M. D. *Electronic Excitation Energy Transfer in Condensed Matter* (North-Holland, Amsterdam, 1982).
- de Gennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, 1979).
- Parker, I. D. *J. appl. Phys.* **75**, 1656–1666 (1994).

ACKNOWLEDGEMENTS. We thank P. Dyreklev and N. Ljungkvist for discussions. This work was supported by the Swedish Research Council for Engineering Sciences (TFR).

Eemian cooling in the Norwegian Sea and North Atlantic ocean preceding continental ice-sheet growth

E. Cortijo*, J. C. Duplessy*, L. Labeyrie*, H. Leclaire*, J. Duprat† & T. C. E. van Weering‡

* Centre des Faibles Radioactivités, Laboratoire mixte CNRS-CEA, 91198 Gif sur Yvette Cedex, France

† Département de Géologie et Océanographie, Université de Bordeaux 1, 33405 Talence, France

‡ Nederlands Instituut voor Onderzoek der Zee, Postbox 59, 1790 AB den Burg, Texel, The Netherlands

CHANGING conditions in the North Atlantic region may drive global climate changes^{1,2}. According to previous reconstructions of the last interglacial (the Eemian), North Atlantic sea surface temperatures (SSTs) were similar to present-day values³. In the Norwegian Sea, even warmer conditions appeared as a single pulse of short duration^{4,5}, whereas the Greenland ice record suggests that the warm interglacial air temperatures were interrupted by several cold periods⁶. Here we use faunal and stable-isotope analyses of foraminifera in two sediment cores from the North Atlantic ocean and Norwegian Sea to reconstruct high-resolution records of SST and sea surface salinity (SSS) during the Eemian interglacial. Our results, which differ significantly from the Greenland record⁶, show a sharp decrease in SST and SSS of the Norwegian Sea, associated with a more moderate cooling and freshening of the North Atlantic at the middle of isotope substage 5e, several millennia before the

beginning of continental ice-sheet growth. Changes in the Norwegian Sea surface conditions appear to have represented an important climate change affecting global atmospheric and thermohaline circulations.

We analysed two sediment cores, V27-60 from the Norwegian Sea and NA87-25 from the North Atlantic on the crest of the Feni ridge⁷. We measured $\delta^{18}\text{O}$ values of benthic foraminifera for stratigraphy and several palaeoclimate proxies for surface waters; planktonic foraminiferal $\delta^{18}\text{O}$, SST and salinity estimates. ($\delta^{18}\text{O}$ is the relative deviation of the $^{18}\text{O}/^{16}\text{O}$ ratio in a sample from that in the PDB standard). Both cores have a high sedimentation rate. Data for the period 145–110 kyr before present (BP) are plotted in Fig. 1. In the Norwegian Sea core V27-60, the glacial–interglacial shift of $\delta^{18}\text{O}$ associated with the transition between isotope stages 6 and 5e (termination II) corresponds to the value of 1.2‰ attributed solely to global ice volume variations⁸. For the same event, the North Atlantic core

NA87-25 records a benthic $\delta^{18}\text{O}$ shift of 2‰. The 0.8‰ excess corresponds to a warming of $\sim 3.5^\circ\text{C}$ around 55°N , reflecting a major reorganization of the deep ocean during the glacial to interglacial transition. Values of $\delta^{13}\text{C}$ have been measured for *Cibicides wuellerstorfi* in both cores ($\delta^{13}\text{C}$ is the relative deviation of the $^{13}\text{C}/^{12}\text{C}$ ratio in a sample from that in the PDB standard). As observed in cores from most oceanic basins⁹, the general trend of the $\delta^{13}\text{C}$ signal is a progressive increase from isotope stage 6 to substage 5d, on which oscillations of smaller amplitude are superimposed.

Summer (July–September average) SST estimates are based on a micropalaeontological transfer function¹⁰ from planktonic foraminiferal assemblages. Core V27-60 indicates a short warm period immediately after the transition between isotope stages 6 and 5, with summer SST close to 7°C , similar to the modern value (Fig. 1a). These surface conditions were a consequence of the thermohaline circulation which carried warm and saline

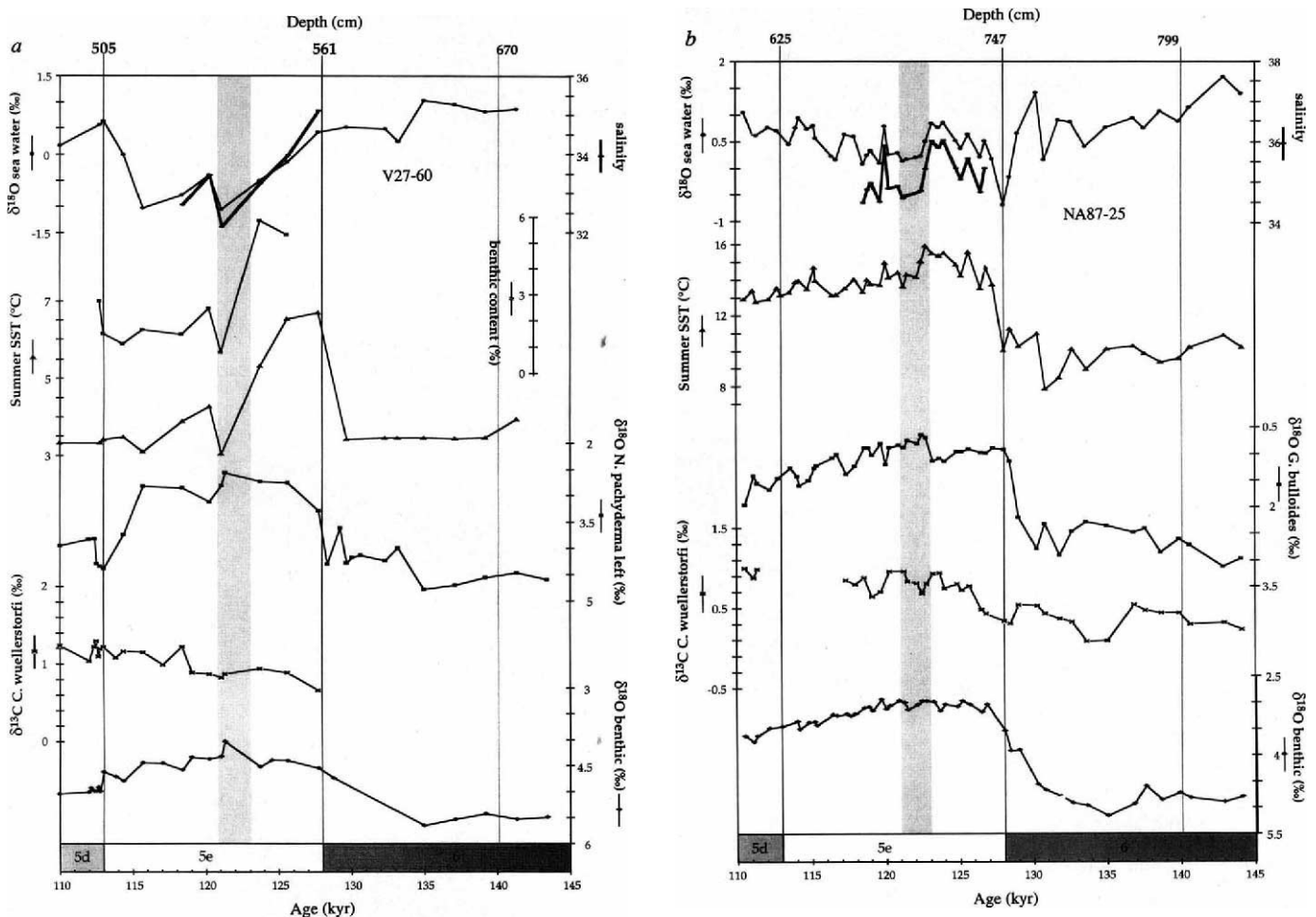


FIG. 1 Data from the two sediment cores. a, Norwegian core V27-60 ($72^\circ 11' \text{N}$, $08^\circ 35' \text{E}$, 2,525 m); b, Atlantic core NA87-25 ($55^\circ 11' \text{N}$, $14^\circ 44' \text{W}$, 2,320 m). Shown from bottom to top are benthic $\delta^{18}\text{O}$, *Cibicides wuellerstorfi* $\delta^{13}\text{C}$, planktonic $\delta^{18}\text{O}$, sea surface temperature (SST), benthic content (for core V27-60), sea surface salinity and sea-water $\delta^{18}\text{O}$ records. (The vertical lines show the transition between stages 6–5e and 5e–5d.)

METHODS. Ages have been estimated by correlating the benthic $\delta^{18}\text{O}$ records with the SPECMAP stack³⁰, giving an age of 128 kyr for the transition between isotope stages 6 and 5e, and 113 kyr for the 5d–5e transition. The shaded vertical bar shows the cooling and freshening of surface water at 123 kyr. The error on this age estimate is ~ 2 kyr, due to the error of the SPECMAP timescale³⁰. The benthic $\delta^{18}\text{O}$ values have been measured on *Cibicides wuellerstorfi* and *Oridorsalis tener* in core V27-60 (ref. 8), and on *C. wuellerstorfi* and *Melonis barleanum* in

core NA87-25. All the isotope values were corrected by a constant factor of +0.64‰ for *C. wuellerstorfi*, +0.37‰ for *O. tener* and +0.40‰ for *M. barleanum* to account for the departure of these species from isotopic equilibrium^{8,31}. At both transition levels (between stages 6–5e and 5e–5d) benthic $\delta^{18}\text{O}$ values are significantly higher than interglacial values, indicating the presence of a significant amount of ice on the continents. In cores V27-60 and NA87-25, the sedimentation rates are 5.3 and 8.1 cm kyr^{-1} , respectively, and our mean sampling resolution is 750 and 500 μm , respectively. The standard error of our SST estimates, which was derived from the statistics on the ten best analogues, is 1.3°C . This is similar to that obtained by calculating SST estimates for each of the 617 core-tops of the Gif data base and comparing them with the modern data. The warming observed during the beginning of substage 5e is therefore significant.

waters towards the high latitudes. The high SST appeared as soon as the continental ice volume ($\delta^{18}\text{O}$ benthic record) reached a minimal value, but they decreased abruptly to 3.5°C (similar to glacial values) in the middle of substage 5e. This cooling was not associated with any significant continental ice volume variation. Assuming a constant sedimentation rate during isotope substage 5e, we estimated that the warm conditions did not last more than 5,000 years and disappeared around 123 kyr ago. This conclusion, which rests on the benthic $\delta^{18}\text{O}$ stratigraphy, differs from that proposed by Kellogg⁵, who used carbonate and planktonic $\delta^{18}\text{O}$ as stratigraphic markers and concluded that warm conditions occurred at the end of isotope substage 5e.

In the North Atlantic, after the penultimate deglaciation, summer SST in core NA87-25 rose to 15.5°C and were 1.7°C warmer than today (Fig. 1b). Subtropical species such as *Globorotalia truncatulinoides* (right coiling) and *Globorotalia hirsuta*¹¹ were then present, indicating that warm waters from the Gulf Stream flowed as far north as 55°N . However, high SSTs did not last throughout the whole of substage 5e. Rather, they decreased sharply by 2°C around 123 kyr BP, at the same time as the major cooling of the Norwegian Sea. They remained close to modern values until substage 5d, when they cooled to 10.5°C . During the transition between isotope substages 5e and 5d, SST changed first slowly and then cooled rapidly after the growth of large amounts of continental ice, in agreement with ref. 12.

We measured planktonic $\delta^{18}\text{O}$ variations of *Neogloboquadrina pachyderma* (left coiling) in core V27-60 and *Globigerina bulloides* in core NA87-25. Using planktonic $\delta^{18}\text{O}$ values and SST estimates, we reconstructed surface water $\delta^{18}\text{O}$ variations during substage 5e by solving the palaeotemperature equation¹³ and assuming that planktonic foraminifera deposited their shells in isotopic equilibrium with the growth temperature, T^* (ref. 14). T^* is (summer SST -1°C) for *G. bulloides* in the range $8-22^\circ\text{C}$, and (summer SST -2.5°C) for *N. pachyderma* (left coiling) in the range $2-10^\circ\text{C}$ (ref. 14). The statistical error is $\pm 0.3\text{‰}$ for $\delta^{18}\text{O}$, and ± 0.6 for salinity estimates. The sea surface $\delta^{18}\text{O}$ variations depend on several parameters: (1) changes in global ice volume, (2) changes in the balance of evaporation, precipitation and runoff, and (3) changes in the rate of advection of saline, high- $\delta^{18}\text{O}$ tropical waters. During the plateau of substage 5e, the continental ice volume was constant and the $\delta^{18}\text{O}$ anomaly (difference from the modern seawater $\delta^{18}\text{O}$) was due only to changes in the precipitation, evaporation, runoff and advection. As long as the ice volume was constant, the local salinity anomaly can be calculated¹⁴, assuming that a seawater $\delta^{18}\text{O}$ anomaly of 1‰ corresponds to a salinity anomaly of 2 (refs 15, 16). In core V27-60, salinities were close to the modern ones (35.2) during the first part of substage 5e. They decreased dramatically to 32.5 simultaneously with the drop of SST, about 123 kyr ago. A similar but smaller change in SSS is observed in core NA87-25: during the beginning of substage 5e, SSSs were slightly higher than today. They reached a maximum value of 36 ± 0.6 , typical of modern Gulf Stream water near 40°N , 40°W during the warmer episode. This indicates that the Gulf Stream flow was significantly larger than today during the first part of the Eemian interglacial. About 123 kyr ago, at the same time as in the Norwegian Sea, SSS decreased rapidly from 36 to 34.7, a value slightly lower than the modern one (35.2). Thus the small cooling and freshening recorded by core NA87-25 coincides in time with the major temperature and salinity drop recorded in core V27-60.

To estimate the representativeness of the SST record of core NA87-25 with respect to the North Atlantic ocean, we used grey reflectance profiles obtained by digitally scanning core photographs and comparing changes in this core with others collected in the $50-56^\circ\text{N}$ latitudinal band (E.C. *et al.*, manuscript in preparation). As the sediment deposited during warm periods is enriched in white coccolith and foraminiferal shells, and exhibits a higher reflectance¹⁷, grey reflectance profiles show changes in the sediment carbonate content; they provide a high temporal

resolution proxy documenting displacements of the polar and subpolar fronts. In all cores, the shape of the grey curve in substage 5e is the same (Fig. 2) showing that they have recorded the same hydrographic conditions. Our data, which agree with the record of core V28-82¹⁷ located close to core SU90-42, indicate that the temperature record of core NA87-25 is representative of a large area of the northeastern Atlantic. We observed no low-reflectance peak similar to that found in core DSDP 609 and which indicates a local input of detrital material during the 6-5e transition (ref. 17, and L.L. *et al.*, manuscript in preparation). This suggests that the North Atlantic experienced a large variability of the thermohaline circulation during the penultimate deglaciation, but not during the Eemian.

Our data are in agreement with European pollen records from France^{18, 20} and show that warm conditions were present in the northeastern Atlantic area during the whole interglaciation. The marine record offers no evidence for the two major cold events depicted in the Greenland ice core⁶ (if such events lasting several millenia had occurred, more than 60% of the SST changes would have been preserved in the sediment record²¹).

The drop in salinity and temperature that occurred in the Norwegian Sea during the middle of the Eemian interglacial (and its slight counterpart in the northeastern Atlantic) are striking phenomena. They can be explained by a decrease in thermohaline circulation due to an input of fresh water, possibly involving melting of ice-sheet and sea ice, an increase of the input of less saline North Pacific surface waters through the Bering Strait, or changes in the balance of precipitation and evaporation. The

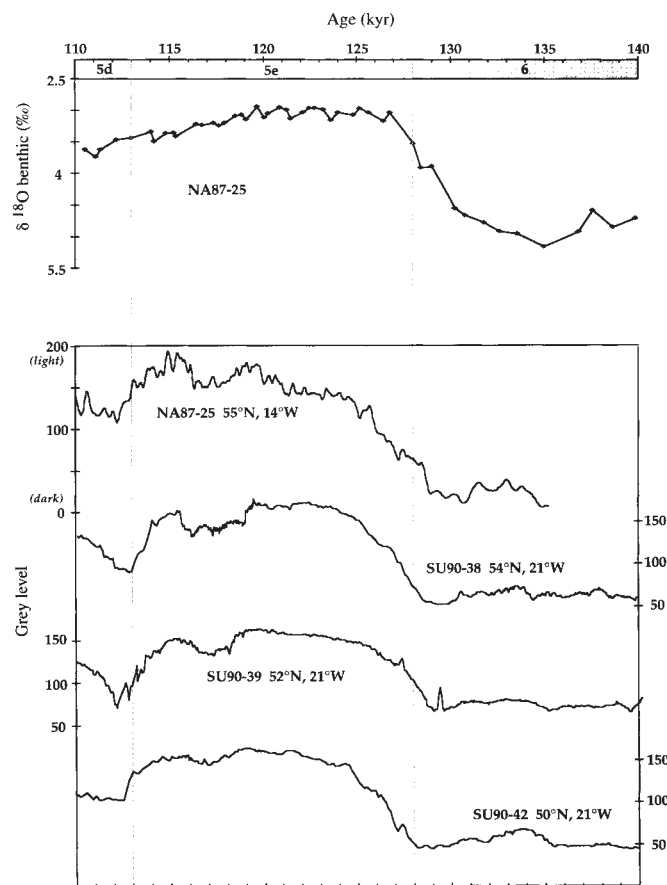


FIG. 2 Grey reflectance for core NA87-25 (top) in substage 5e compared to other North Atlantic cores (bottom). The high numbers indicate a high-reflectance (clear) sediment and the low numbers show a low-reflectance (dark) sediment. The benthic $\delta^{18}\text{O}$ is shown for the stratigraphy.

ice volume was minimal during the substage 5e, and the plateau in the benthic $\delta^{18}\text{O}$ record demonstrates that the continental ice volume did not experience significant changes. The quantity of melting ice was thus small. It was the same for icebergs and meltwater fluxes. The absence of detrital material in Norwegian Sea core V27-60 (associated with the temperature and salinity drops) confirms that no massive iceberg discharges occurred 123 kyr ago.

Two characteristics of the Eemian climate favoured instabilities in the freshwater balance of the Arctic and Norwegian seas. (1) As the mean sea level was higher than today by 4–7 m (ref. 22), a larger flow of Pacific surface waters could cross the Bering Strait, bringing cold fresh water to the Norwegian Sea²³. Even during the warmer period of substage 5e, the distribution of subpolar fauna indicates that cold water flowed to the southeast off the northern coast of Iceland⁵. The flux of Pacific water would largely depend on changes in the atmospheric circulation and would be affected by its variability²³. (2) High SST resulted in strong evaporation, leading to increased precipitation at high latitudes (Arctic seas and surrounding continents)²⁴. These two sources of fresh water drive both an increased formation of sea ice^{25,26} and an increased runoff input to the Arctic Sea. Accordingly, the flow of exported fresh water towards the Norwegian Sea would be larger. However, Arctic sea ice was not exported to the location of core V27-60 (around 72° N) because this core contains calcareous nanofossils during the whole stage 5 (ref. 27).

Model simulations indicate that a slight decrease in surface water salinity could have slackened (or stopped) deep-water formation in the Norwegian Sea within a few decades²⁸. The thermohaline circulation and the deep convection would then be affected, as shown by the reduction of the total benthic content in core V27-60 (Fig. 1a) and by the 0.2‰ decrease in *Cibicides* $\delta^{13}\text{C}$ in both cores. As the Norwegian Sea deep water is very homogenous below the depth of the sill which separates the deep Norwegian basin from the Atlantic (about 600 m), our data indicate that the ventilation of the whole deep Norwegian Sea and the eastern Atlantic at 2,300 m depth was significantly reduced when both SST and SSS decreased. This was sufficient to produce a significant cooling of the Norwegian Sea, strengthening and enhancing the SST gradient between this water mass and the North Atlantic ocean. Such a pattern would favour the formation of major atmospheric depressions, bringing winter snow and promoting ice-sheet growth over the continents. Our data, when put together with the results of Duplessy and Shackleton⁹, show that the rate of deep-water formation in the Norwegian Sea decreased 123 kyr ago, but did not cease completely. Nevertheless, the North Atlantic Deep Water flow was still very active^{9,29}. The climatic evolution that we have reconstructed supports the scheme proposed by Imbrie *et al.*^{1,2} to explain the phasing of the ocean and climate response to changes in Northern Hemisphere insolation: the earliest response to such high-latitude changes is a decrease of thermohaline circulation in the Norwegian Sea and a subsequent transfer of the main deep-water convection to the North Atlantic. □

13. Shackleton, N. J. Attainment of Isotopic Equilibrium between Ocean Water and Benthonic Foraminifera Genus *Uvigerina*: Isotopic Changes in the Ocean during the last Glacial 203–209 (Rep. No. 219, CNRS, Paris, 1974).
14. Duplessy, J. C. *et al.* *Oceanologica Acta* **14**, 311–324 (1991).
15. Craig, H. & Gordon, L. I. in *Stable Isotopes in Oceanographic Studies and Paleotemperatures* (ed. Tongiorgi, E.) 9–122 (CNR, Spoleto, 1965).
16. Lewis, E. L. & Perkin, R. G. *J. geophys. Res.* **83**, 466–478 (1978).
17. Bond, G., Broecker, W., Lotti, R. & MacManus, J. in *Start of a Glacial* (eds Kukla, G. J. & Went, E.) 185–205 (NATO ASI Ser. 13, Springer, Berlin, 1992).
18. Mangerud, J., Sonstegaard, E. & Sejrup, H. P. *Nature* **277**, 189–192 (1979).
19. Guiot, J., Pons, A., de Beaulieu, J. L. & Reille, M. *Nature* **338**, 309–313 (1989).
20. Woillard, G. *Nature* **281**, 558–562 (1979).
21. Bard, E., Arnold, M., Duprat, J., Moyes, J. & Duplessy, J. C. *Clim. Dyn.* **1**, 101–112 (1987).
22. Lambeck, K. & Nakada, M. *Nature* **357**, 125–128 (1992).
23. Shaffer, G. & Bendtsen, J. *Nature* **367**, 354–357 (1994).
24. Manabe, S. & Stouffer, R. J. *Nature* **364**, 215–218 (1993).
25. Aagaard, K. & Carmack, E. C. *J. Geophys. Res.* **94**, 14485–14498 (1989).
26. Mysak, L. A., Manak, D. K. & Marsden, R. F. *Clim. Dyn.* **5**, 111–133 (1990).
27. Gard, G. *Quat. Sci. Rev.* **7**, 65–78 (1988).
28. Bryan, F. *Nature* **323**, 301–304 (1986).
29. Duplessy, J. C. *et al.* *Quat. Res.* **21**, 225–243 (1984).
30. Imbrie, J. *et al.* in *Milankovitch and Climate* (eds Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (NATO ASI Ser. C126, Reidel, Dordrecht, 1984).
31. Duplessy, J. C., Moyes, J. & Pujol, C. *Nature* **286**, 479–482 (1980).

ACKNOWLEDGEMENTS. We thank D. Anderson, F. Bassinot and L. W. Gates for reviews of the manuscript, U. Pflaumann for providing foraminiferal counts made at Kiel University, and J. Antignac, B. Le Coat and J. Tessier for processing of the isotope analyses. This work was supported by CNRS, CEA, INSU (PNEDC) and EU Environment Programme. T.C.E. van W. was supported by NIOZ and SOZ.

Basalt vesicularity as a measure of atmospheric pressure and palaeoelevation

Dork L. Sahagian* & Joseph E. Maus†

* Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA
† Department of Geological Sciences, The Ohio State University, Columbus, Ohio 43210, USA

THE pressure in, and thus the size of, a bubble in a lava flow is determined by the atmospheric pressure and the hydrostatic pressure of the overlying lava. If atmospheric sea-level pressure is known (or assumed), vesicle size distributions in basalt flows can thus be used as an indicator of the palaeoelevation of emplacement¹. Here we show by analysis of the vesicle size distribution of basalt samples collected from the summit and base of Mauna Loa volcano in Hawaii that the technique provides estimates of ambient pressure that are accurate to within 0.1 bar, and thus estimates of elevation with a resolution of about 1,400 m. This palaeoaltimetric technique should find useful application in palaeogeographical reconstruction² for times when sea-level pressure was similar to the present value. (Conversely, if the elevation of emplacement is known, our technique should be able to provide a measure of sea-level palaeopressure.) Unlike palaeobotanical methods for estimating elevations^{3–6}, this method does not have to assume a relationship between temperature and altitude; on the other hand, its use is restricted to lava flows that have had a simple emplacement history.

When a lava is emplaced, bubbles containing equal amounts of gas at the base and top of a flow will be subject to different total pressures because of difference in overburden. The atmospheric-pressure dependence of vesicle size can be expressed by the following ratio of vesicle distributions at the top and bottom of a flow:

$$\frac{V_t}{V_b} = \frac{P + \rho g H}{P} \quad (1)$$

where V_t and V_b are the volumes of the modal bubble sizes at the top and bottom of the flow, respectively, ρ is lava density, g is the acceleration due to gravity, H is flow thickness and P is atmospheric pressure at emplacement. We use modal instead of

Received 6 May; accepted 18 October 1994.

1. Imbrie, J. *et al.* *Paleoceanography* **7**, 701–738 (1992).
2. Imbrie, J. *et al.* *Paleoceanography* **8**, 699–735 (1993).
3. CLIMAP Project Members *Quat. Res.* **21**, 123–224 (1984).
4. Duplessy, J. C. & Labeyrie, L. in *Start of a Glacial* (eds Kukla, G. J. & Went, E.) 173–183 (NATO ASI Ser. 13, Springer, Berlin, 1992).
5. Kellogg, T. B. *Boreas* **9**, 115–137 (1980).
6. GRIP Project Members *Nature* **364**, 203–207 (1993).
7. Van Weering, T. C. E. & Rijk, S. D. *Mar. Geol.* **101**, 49–69 (1991).
8. Labeyrie, L., Duplessy, J. C. & Blanc, P. L. *Nature* **327**, 477–482 (1987).
9. Duplessy, J. C. & Shackleton, N. J. *Nature* **316**, 500–507 (1985).
10. Prell, W. L. *The Stability of Low-latitude Sea-surface Temperatures: an Evolution of the Climap Reconstruction with Emphasis on the Positive SST Anomalies* (Rep. No. TR025, US Department of Energy, Washington DC, 1985).
11. Be, A. W. H. & Tolderlund, D. S. in *The Micropaleontology of the Oceans* (eds Funnel, B. M. & Riedel, W. R.) 105–149 (Cambridge Univ. Press, 1971).
12. Ruddiman, W. F. & McIntyre, A. *Science* **204**, 173–175 (1979).

average size because it is more easily measured and less sensitive to the presence or absence (or loss) of a few large vesicles.

The application of this simple approach to natural lavas for the purpose of determining atmospheric pressure requires the following assumptions to be made. (1) Basaltic lava was extruded with a well mixed population of bubbles so that their initial mass distribution is not a function of vertical position in the flow. (2) The lava flow experienced a simple emplacement history. (3) The flow was erupted from a volcanic vent, travelled to its final resting place in a short time (a few hours) and solidified in place without further disturbance. Such disturbance could include re-inflation by additional lava after top and base had solidified, deflation by drainage of interior lava, and additional lavas emplaced over the flow causing the solidified top to shear away from the site of emplacement. (4) There are no sources of bubbles, such as soil moisture or burning vegetation, outside the flow to introduce bubbles into the flow before solidification of top and base.

Samples were collected from Mauna Loa, Hawaii, to test the sensitivity of our technique. Flows were sampled only where a complete section from top to base was exposed. Low-elevation samples were taken from the vicinity of Hilo, whereas high-elevation samples were taken from the peak of Mauna Loa, on the east rim of the caldera. We collected samples from parts of the flows which appeared to have experienced a simple emplacement history.

Coalesced bubbles were not observed until at least several centimetres in from the top and base of the flows. This is attributed to rapid cooling of the top and bottom surfaces. We were careful to avoid overturned or rubbly flow tops, flows which

displayed significant lateral variation of thickness, sites in which the flow overran an obstacle, and injection of volatiles from burning trees, surface and ground water. We sampled vesicular sheet flows⁷ of uniform thickness which displayed no evidence of being part of a lava tube system. The flow tops were vesicular and lacked a dense glassy rind.

We used three criteria for selecting sites with simple emplacement histories. (1) Uniformity of flow. There should be no local variations in flow thickness, which is simple to test in the field. (2) Location of samples. A lobe away from the main lava flow is unlikely to have deflated, although inflation is still possible. This criterion is both more difficult to test in practice, and less robust. (3) Bubble space and size distribution. Modelling of bubbles in cooling lavas (see, for example, refs 1 and 21) has shown that late-stage inflation (or deflation) can be recognized by its influence on the final space and size distribution of the bubbles. This criterion is the most robust, but requires careful examination in the field. (In the present work, we made further selection on the basis of laboratory analyses of bubble size and spatial distributions in samples collected from each flow.) Although none of these criteria can completely eliminate the possibility of inflation or deflation after solidification of the upper and lower 10 cm of a flow, they provide guidelines for selecting flows that have the greatest likelihood of providing reliable data.

Samples were analysed in the laboratory to determine vesicle size distributions by impregnating the samples with a plastic polymer, dissolving the basalt in hydrofluoric acid, and measuring the plastic casts of the vesicles¹. To determine a representative size distribution, samples were several centimetres in each

FIG. 1 Observed vesicle size distributions at the top and bottom of two flows, one at 50 m and the other at 3,915 m above sea level. Vesicles obtained from basalt samples were sorted into size groups with increments of 1 mm. For comparison, a Poisson distribution was constructed with the same size resolution. Note that the modes match very closely. It is important to distinguish the positions of the modes from those of the means of the distributions. The amplitude is not used in this analysis, as it is subject to the vagaries of emplacement and degassing. Although the laboratory technique generated three-dimensional vesicle casts for exact size distribution determination, the difficulty and tedium of sorting and counting many thousands of casts made it impractical to use a finer size grouping than 1 mm.

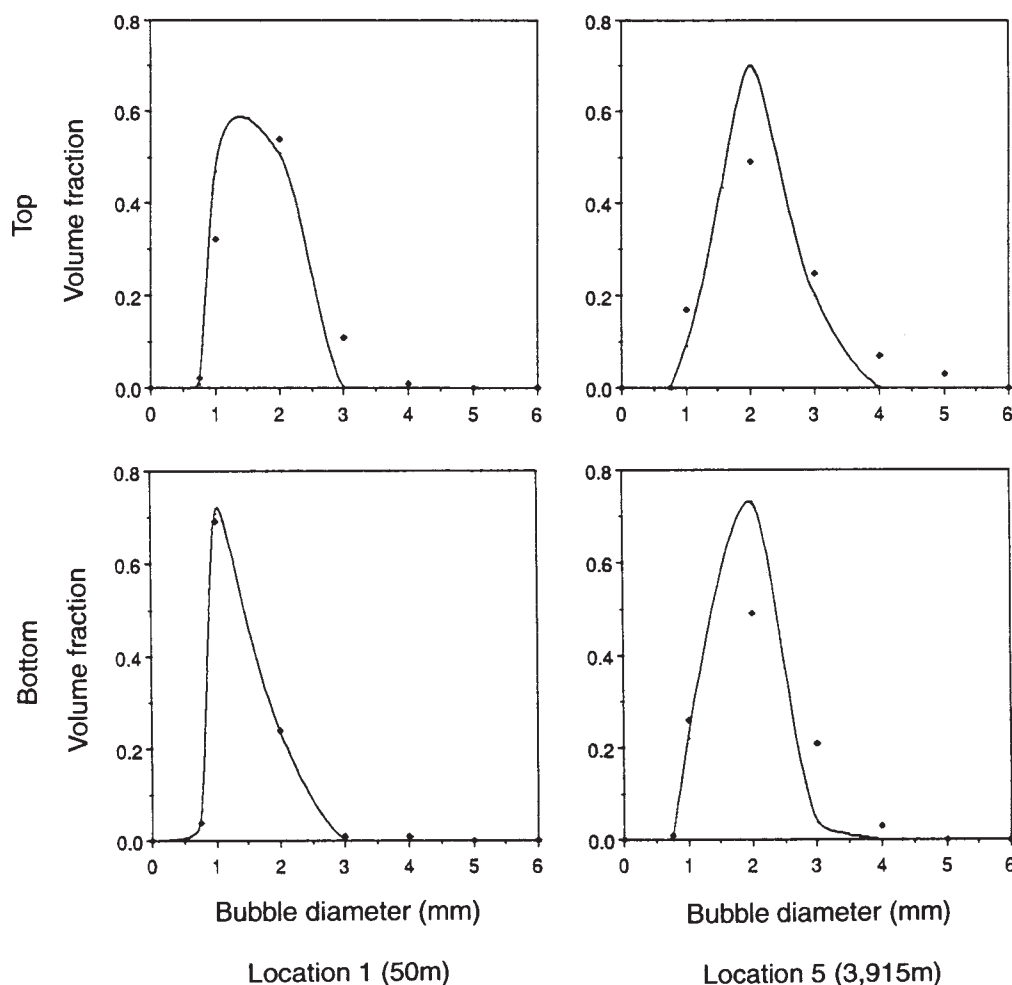


TABLE 1 Atmospheric pressure based on observed basalt vesicularities

Location	Elevation (m)	Flow thickness (m)	Observed modal volume ratio	Inferred pressure (bar)	Standard atm. pressure (bar)
1	50	1.52	1.545	0.89	1.0
3	300	1.78	1.585	0.92	0.99
5	3,915	1.02	1.585	0.54	0.63
6	3,915	1.23	1.581	0.57	0.63
7	3,870	1.65	1.807	0.57	0.64
8	3,870	1.27	1.685	0.56	0.64
9	3,909	1.68	1.953	0.54	0.63

Observed modal volume ratios were obtained by interpolating measured size distributions. This allowed modes to be determined even when they had fractional-millimetre values. If extremely fine resolution sampling were possible, one could obtain modal size directly from basalt samples. However, very fine gradations of measured bubble sizes would be necessary in order to determine a mode without interpolation. Our simple interpolation makes it possible to determine the fractional modal size. Inferred pressure was obtained by using modal ratios in equation (1) and solving for P . Agreement of results with the standard atmosphere indicates accuracy within ~ 0.1 bar. The sensitivity of the technique is maximized for lava flow of thickness such that $pg_h = 1$ atm (h is flow thickness). At sea level, this thickness is 3 m. Sampled flow thicknesses ranged from 1.02 to 1.68 m, so it is expected that the resolution of the results would have been better if thicker flows had been available.

dimension, containing thousands of vesicles. After impregnation, samples were cut in half, revealing that penetration was complete for every sample. The size distribution of vesicles was quantified by counting and sorting the plastic casts into size categories with 1 mm increments of diameter. The casts were sorted as if they were perfect spheres. Coalescence of bubbles was observed in samples from the interior of flows, but not from top and bottom, consistent with field observations. The plastic casts were resistant to the effects of hydrofluoric acid, so they reliably represented the actual vesicle sizes. The arduous nature of the laboratory analysis suggests that future studies should use two-dimensional statistical techniques^{8–10}. Such techniques analyse two-dimensional sample sections to estimate maximum, minimum, mean and mode sizes. The rapid analysis would permit larger (and more) samples to be counted, thereby reducing the potential effects of local vesicularity variations, or biases caused by presence or lack of individual vesicles of the largest size.

To determine the modal size as precisely as possible, a Poisson size distribution was constructed¹ with 0.01-mm resolution, and degraded (grouped) to match the 1-mm resolution of the observations (Fig. 1). The full resolution of the Poisson distribution was then used to interpolate the observed size distribution. A Poisson distribution of diameter was used (rather than gaussian or other distributions) because the majority of vesicles have been observed to be smaller than the mode. If extremely fine resolution sampling were possible, one could obtain modal size directly from natural samples without interpolation. Observed resolution of 0.1 mm should be possible with more sophisticated measuring techniques, with interpolation possible to 0.001 mm. However, the resolution of this technique, when applied to the estimation of atmospheric pressure, will still be limited by the validity of the assumptions (1)–(4) described above. Although the sort-

ing interval of 1 mm used for samples collected from Mauna Loa was sufficient to test our technique (Fig. 1), future applications would be enhanced by a finer sorting interval.

The modes were used to calculate atmospheric pressure for each observed flow from equation (1). Table 1 gives the ratio of modal size between top and base of each flow. The inferred pressures agree well with those of the standard atmosphere as calibrated by measurements at Hawaii. The resolution of our results is ~ 0.1 bar, which corresponds to an elevation resolution of 1–1.4 km.

The inferred atmospheric pressures (Table 1) are consistently lower than the standard atmosphere at the elevations from which samples were collected. Although the magnitudes of the discrepancies are not large, they warrant some re-examination of our assumption of simple emplacement history. The following variations would serve to reduce the inferred atmospheric pressure. Deflation of the lava flows after solidification of the tops and bases would cause the measured (lesser) thickness to produce an inferred pressure less than the actual value (equation (1)). If the solidifying crust allowed a hydrostatic head at the flow front, the internal lava pressure would have been greater than the sum of atmospheric pressure and vertical lava overburden⁷. Further, if the top of a flow solidified to depth of more than 10 cm while still in motion over the fluid interior, the vesicles in the upper crust may have recorded a lower pressure (higher elevation) than appropriate for the site of final emplacement.

A source of error in our investigation was the sorting stage of vesicle counting. Transitional sizes were placed as consistently as possible into larger or smaller categories, but considerable random or systematic error could have been introduced, particularly in the division between 1 and 2 mm. For example, if 20% of the 2-mm-diameter casts were erroneously categorized as 1 mm, the mode would shift up by 0.004 mm (for location 5 base), which corresponds to a downward shift of calculated atmospheric pressure of 3%. This could account for as much as half of some of the discrepancies in Table 1. Despite these possible effects, the results fall within a narrow range about 0.1 bar less than the standard atmosphere at the given elevation.

Whereas palaeoelevation techniques based on floral analysis cannot be used above the tree line, the vesicle-size technique should be particularly useful for highland areas, where basalts are often emplaced and for which there are few alternative methods for estimating palaeoelevation^{11,12}. Data of this sort would be useful for constraining the timing of epeirogenic activity. For instance, if suitable basalt flows on the Tibetan Plateau¹³ or Colorado Plateau¹⁴ can be identified, sampled and dated, it will be possible to determine the age of uplift of the various parts of the plateaus, perhaps leading to insights regarding uplift mechanisms¹⁵ and the relationship between climate and orogeny¹⁶.

Conversely, if the elevation of flow emplacement is known (as has been argued for the Keweenaw lavas, for example) atmospheric sea-level pressure can be calculated for times in the past. Although Phanerozoic sea-level pressure has probably not varied significantly, there are no reliable data that bear on Precambrian sea-level pressure, although there are many data constraining past atmospheric composition^{17–20}. Such measures of sea-level pressure in the distant past might be used to constrain models of planetary degassing. □

Received 25 January; accepted 4 October 1994.

- Sahagian, D. L., Anderson, A. T. & Ward, B. *Bull. volcan.* **52**, 49–56 (1989).
- Gregory, K. & Chase, C. *Geology* **20**, 581–585 (1992).
- Meyer, H. W. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **99**, 71–99 (1992).
- Wolfe, J. & Shorn, H. *Paleobiology* **15**, 180–198 (1989).
- Wolfe, J. *Bull. U.S. Geol. Surv.* **1964**, 1–35 (1992).
- Ziegler, A. M. et al. *Rev. Earth planet. Sci.* **13**, 385–425 (1985).
- Hon, K., Kauahikaua, J., Denlinger, R. & Mackay, K. *Bull. geol. Soc. Am.* **106**, 351–370 (1994).
- Mangan, M. T., Cashman, K. V. & Newman, S. *Geology* **21**, 157–160 (1993).
- Cashman, M., Mangan, M. & Newman, S. (abstr.) *EOS* **73**, 629 (1992).
- Pingatore, N. et al. (abstr.) *EOS* **74**, 626 (1993).
- Sahagian, D. L. *J. geophys. Res.* **92**, 4895–4904 (1987).

- Sahagian, D. L. *Tectonics* **7**, 125–138 (1988).
- Turner, S. et al. *Nature* **364**, 50–54 (1993).
- Lipman, P. W. & Mehnert, H. H. *Mem. geol. Soc. Am.* **144**, 119–154 (1975).
- Beghoul, N., Barazangi, M. & Isacks, B. *J. geophys. Res.* **98**, 1997–2016 (1993).
- Molnar, P. & England, P. *Nature* **346**, 29–34 (1990).
- Barron, E. J. & Washington, W. M. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. et al.) 546–553 (Am. Geophysical Union, Washington DC, 1985).
- Abe, Y. & Matsui, T. *J. geophys. Res.* **91**, E291–E302 (1986).
- Hansen, J. et al. *Science* **213**, 957–966 (1981).
- Sundquist, E. T. & Broecker, W. S. *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* 1–627 (Am. Geophysical Union, Washington DC, 1985).
- Sahagian, D. L. *J. Geol.* **93**, 205–211 (1985).

Prediction of crystal–melt partition coefficients from elastic moduli

Jon Blundy & Bernard Wood

CETSEI, Department of Geology, Bristol University,
Wills Memorial Building, Bristol BS8 1RJ, UK

MANY geochemical processes, such as crystallization of silicate magmas or planetary differentiation, require a knowledge of the way in which elements become partitioned between coexisting crystal and liquid phases^{1,2}. But quantitative prediction of crystal/melt partition coefficients from thermodynamic principles has not previously been possible. By studying the partitioning of 15 elements between silicate minerals and their coexisting melts, we show here that the partitioning behaviour of any series of isovalent cations can be rationalized in terms of a simple model in which the size and elasticity of the crystal lattice sites play a critical role. We find that elasticity varies linearly with the formal charge of the cation. This model allows us to predict element partitioning behaviour solely from the physical characteristics of the cation sites in the crystal.

Partition coefficients may vary by several orders of magnitude, even for the case of a single element entering a single mineral phase, as the conditions of crystallization are varied^{1,2}. As the Nernst partition coefficient (D_i , defined as $[i]_{\text{mineral}}/[i]_{\text{melt}}$)³ is related to the equilibrium constant for an appropriate exchange or fusion reaction^{4,5}, much of this variation can be attributed to the combined effects of pressure (P), temperature (T) and phase composition. The importance of these variables is widely recognized and yet the quantitative links between them and partition coefficients are poorly understood. Consequently, although most processes of chemical differentiation occur under conditions of varying P , T and composition, a lack of understanding of their effects obliges geochemists to adopt a largely empirical approach using constant average D_i values in modelling.

To a first approximation, partitioning energetics can be divided into two parts, $\Delta G_{\text{fusion}}^y$ and $\Delta G_{\text{exchange}}^{y-i}$, such that

$$D_i = \exp \left(\frac{\Delta G_{\text{fusion}}^y - \Delta G_{\text{exchange}}^{y-i}}{RT} \right) \quad (1)$$

where R is the gas constant and T is in degrees K. $\Delta G_{\text{fusion}}^y$ is the standard-state free energy of fusion of the host mineral. In multicomponent systems, $\Delta G_{\text{fusion}}^y$ governs mineral/melt partitioning of the host cation (y) onto the lattice site of interest as a function of P and T . $\Delta G_{\text{exchange}}^{y-i}$ is the free energy required to remove from the melt and insert into the crystal lattice a cation, i , differing in size and/or charge from y , and to simultaneously transfer y from crystal to melt. As minerals are typically less compressible than coexisting melts, $\Delta G_{\text{exchange}}^{y-i}$ is dominated by the strain energy around the mismatched (or defect) cation in the crystal^{6–11}, with melt structure playing a subordinate role⁹.

Brice⁸ presents an expression relating the mechanical strain energy around a homovalent cation defect (ΔG_{strain}), to the Young's Modulus (E) of the host crystal:

$$\Delta G_{\text{strain}} = 4\pi EN_A \left[\frac{r_0}{2} (r_i - r_0)^2 + \frac{1}{3} (r_i - r_0)^3 \right] \approx \Delta G_{\text{exchange}}^{y-i}$$

where r_0 is the optimum radius of the lattice site, r_i is the radius of the substituent cation and N_A is Avogadro's number. Replacing $\Delta G_{\text{fusion}}^y$ in equation (1) with a 'strain-compensated partition coefficient', $D_0(P, T, X)$, that describes strain-free cation substitution ($r_i = r_0$) at the P , T and composition of interest, we obtain

an expression relating $D_i(P, T, X)$ to $D_0(P, T, X)$, r_0 and E :

$$D_i(P, T, X) = D_0(P, T, X) \times \exp \left[\frac{-4\pi EN_A \left[\frac{r_0}{2} (r_i - r_0)^2 + \frac{1}{3} (r_i - r_0)^3 \right]}{RT} \right] \quad (2)$$

The relationship between $D_0(P, T, X)$ values for different cation valences (D_0^{2+} , D_0^{3+} and so on) should depend upon the

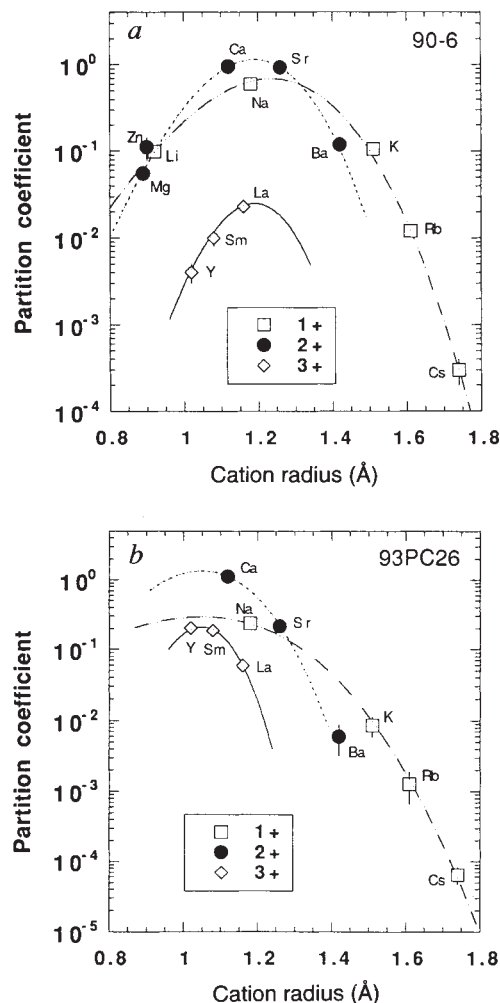


FIG. 1 Plots of experimentally determined partition coefficient versus cation radius for a, plagioclase (An_{89}) from run 90-6 (1 atm, 1,251 °C); and b, clinopyroxene (Di_{87}) from run 93PC26 (30 kbar, 1,665 °C). Only data for elements partitioning into the plagioclase M site and clinopyroxene M2 site are shown. Ionic radii are for eight-fold coordination in both minerals³¹. Curves drawn through series of isovalent cations are obtained by weighted non-linear least-squares regression to equation (2). Fit parameters are given in Table 1; 1 s.d. error bars, based on at least three replicate analyses, are only shown where they exceed the size of the plot symbol. Experiments were done at the University of Bristol using vertical quench furnaces (1 atm) and piston cylinder apparatus (10–30 kbar) with BaCO_3 pressure cells. Temperature is believed to be accurate to within ± 5 °C; pressure (after $\sim 10\%$ pressure correction) to within ± 0.5 kbar. Starting materials were glassed-gel mixtures doped with p.p.m. levels of Li, K, Rb, Cs, Sr, Ba, Zn, Ga, Y, La, Sm and Pb, and welded into Pt capsules. Quenched run products were analysed by wavelength-dispersive electron microprobe analysis (EMP; University of Bristol) and secondary ion mass spectrometry (SIMS; University of Edinburgh)³². Accuracy, based on analysis of synthetic secondary standards, is better than $\pm 2\%$ relative for EMP and $\pm 15\%$ relative for SIMS for all elements except Ga and Zn ($\pm 30\%$), which suffer substantial molecular interference.

TABLE 1 Comparison of lattice site parameters estimated from partitioning data, with bulk crystal properties

	Charge	Parameter	Plagioclase (run 90-6)	Diopside (run 93PC26)	Augite (HD)*	Anorthite (mean)†	Albite (mean)†	Diopside (mean)‡
Lattice site§	1+	$\bar{E}$ (kbar)	644 (31)	504 (8)	—	673 (70)	516 (39)	404 (4)
		r_0 (Å)	1.235 (5)	[1.05]	—	1.216 (31)	1.237 (16)	[1.05]
	2+	$\bar{E}$ (kbar)	1,116 (35)	1,465 (89)	1,608 (61)	1,245 (148)	1,044 (82)	1,819 (30)
		r_0 (Å)	1.190 (2)	[1.05]	[1.02]	1.196 (18)	1.290 (11)	[1.05]
	3+	$\bar{E}$ (kbar)	1,877 (103)	3,956 (311)	3,535 (174)	1,903 (213)	2,084 (119)	3,581 (49)
		r_0 (Å)	[1.190]	[1.05]	1.02 (1)	[1.196]	[1.290]	[1.05]
Bulk crystal¶		E (kbar)	—	—	—	1,074 (40)	622 (40)	1,680 (4)
		$d-1.38$ (Å)#	—	—	—	1.15 (17)	1.22 (20)	1.12 (16)

* Regression results from experimental data of Hart and Dunn¹⁸. Clinopyroxene is Al, Ti-rich sub-calcic augite equilibrated with basalt melt at 30 kbar, 1,380 °C.

† Average values for anorthite and albite are obtained by linear extrapolation of plagioclase data (17 experiments) to end-member compositions.

‡ Average values for diopside (7 experiments).

§ Values for lattice sites M (plagioclase) and M2 (diopside) obtained by regression of experimental data. 1 s.d. quoted in parentheses in terms of last significant digits.

* Bulk crystal data from refs 9, 21, 26.

Estimated optimum cation radius (r_0) for plagioclase M site or clinopyroxene M2 site, calculated as mean cation–oxygen distance¹⁷ less the ionic radius of O²⁻ (1.38 Å)³¹.

|| Values in square brackets were imposed on the regressions in cases where only three D_i values per experiment were available.

configurational entropy of the doped crystal, which for heterovalent substitution depends upon the charge-balancing mechanism. In the case of homovalent substitution, $D_0(P, T, X)$ should approximate $D_i(P, T, X)$ for the site of interest; that is, $r_i \approx r_0$. Thus for the M2 site of diopside ($\text{CaMgSi}_2\text{O}_6$) $D_0^{2+}(P, T, X) \approx D_{\text{Ca}}(P, T, X)$.

In order to test this simple theory, we have performed mineral-melt partitioning experiments for clinopyroxene and plagioclase over the pressure range 1 bar to 30 kbar, in the simple system $\text{CaMgSi}_2\text{O}_6\text{--NaAlSi}_3\text{O}_8\text{--CaAl}_2\text{Si}_2\text{O}_8$ doped with p.p.m. levels of 12 elements of variable cation charge and size. Representative results and brief experimental and analytical details are given in Fig. 1 (to be presented in full elsewhere).

It is well known^{2,12,13} that plots of $\log D_i$ versus r_i for series of isovalent cations describe parabolas with maxima corresponding to the size of the crystal lattice site(s) on which substitution occurs (Fig. 1). Heats of solution of homovalent impurities in oxides, as determined computationally, show a similar parabolic dependence on cation size^{14,15}. Plagioclase has two types of cation sites; clinopyroxene has three. Here we confine ourselves to the large plagioclase M sites and clinopyroxene M2 site onto which the majority of the elements that we studied partition. These sites are distorted with ~ 8 -fold coordination. (Note that the adoption of 6- or 7-fold coordination for plagioclase M sites¹⁶ has no material effect on our results.) By fitting the experimentally determined D_i values to equation (2), we can obtain best-fit values for $D_0(P, T, X)$, $\bar{E}$ and r_0 for each isovalent series where the bar denotes a site property rather than a bulk-crystal property (Table 1).

For 1⁺ and 2⁺ cations in plagioclase, we find that r_0 increases from anorthite ($\text{CaAl}_2\text{Si}_2\text{O}_8$) to albite ($\text{NaAlSi}_3\text{O}_8$) (Table 1), consistent with the increase in mean M–O bond length along the solid solution series^{16,17}. As only three trivalent cations (Y, Sm, La) partition onto the plagioclase M site in this study, we have set $r_0^{3+} = r_0^{2+}$. Similarly, for clinopyroxene we are obliged to fix r_0 for all valences: our chosen value (1.05 Å) is based on the observation that $D_{\text{Sm}} \approx D_{\text{Y}}$ (Fig. 1b), and is in keeping with known M2–O bond lengths in diopside¹⁷ (Table 1).

Best-fit values of $\bar{E}_{\text{plag}}$ depend on plagioclase composition and, to a greater extent, on the charge of the substituent cation (Fig. 2 and Table 1). For clinopyroxene, $\bar{E}_{\text{cpx}}$ varies similarly with charge, both for the new experimental data and for those of Hart and Dunn¹⁸ (Table 1). Significantly, they find $D_{\text{Y}} > D_{\text{Sm}}$, resulting in a value of r_0 (1.02 $\pm$ 0.01 Å) that is smaller than ours (1.05 Å). This is in keeping with the higher Mg/Ca ratio of their clinopyroxene and the shorter M2–O bond length in enstatite (MgSiO_3) relative to diopside¹⁷. Fits of other experimental rare

earth element (REE) partitioning data¹⁹ also yield $r_0 \approx 1.02$ Å, suggesting that for typical magmatic sub-calcic augites this value is most appropriate. It is possible that the apparent pressure dependence of r_0 in clinopyroxene¹⁹ is in fact primarily a consequence of changing enstatite content (and hence M2–O bond length).

Two features of our results are significant. First, for homovalent substitution, $\bar{E}$ is similar in magnitude to E for the bulk crystal (Fig. 2 and Table 1). This may be coincidence, but it could also imply that the elastic properties of the crystal are strongly influenced by those of the large cation site, rather than by the relatively rigid tetrahedral Si–O(–Al) framework^{20,21}. Second, $\bar{E}$ is an approximately linear function of cation charge (Fig. 2 and Table 1). This behaviour is analogous to that of ionic crystals, wherein for a given structure the bulk modulus, K , is a function of the ratio of cation charge (z_c) to molecular volume²². Hazen and Finger²³ have observed a similar relationship between bulk moduli of cation–oxygen polyhedra and the ratio of z_c to polyhedral volume in a wide variety of minerals and oxides. Values of $\bar{E}$ obtained from partitioning data (Table 1) can be converted to site bulk moduli ($\bar{K}$) via Poisson's ratio, which is taken to be 0.25; that is, we assume that the site behaves as a perfect Poisson solid, in which the Lamé constants λ and

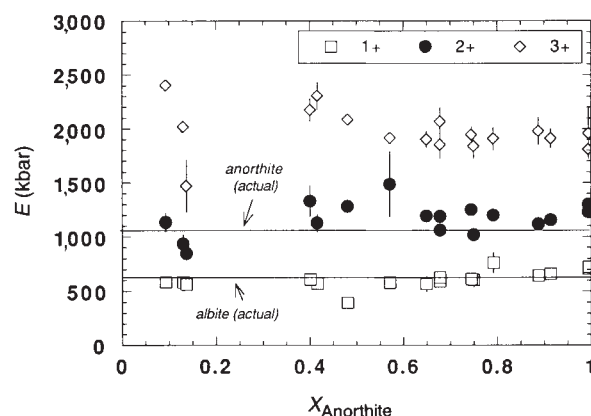


FIG. 2 Variation in lattice-site Young's Modulus ($\bar{E}$) with anorthite content of the host crystal for the plagioclase partitioning experiments. Horizontal lines labelled 'anorthite' and 'albite' denote measured Young's Moduli of the bulk crystals²¹. Note the similarity between $\bar{E}_{\text{plag}}^{1+}$ and $\bar{E}_{\text{albite}}^{1+}$, and $\bar{E}_{\text{plag}}^{2+}$ and $\bar{E}_{\text{anorthite}}^{2+}$. Mean values for $\bar{E}_{\text{plag}}^{1+}$ and $\bar{E}_{\text{plag}}^{2+}$ from the experimental data are given in Table 1. Error bars are 1 s.d.

μ are equal. Our values of $\bar{K}$ are in good agreement with those from polyhedral data²³ (Fig. 3), which is remarkable as the former were obtained solely from chemical partitioning data.

The quantitative relationship between the elastic properties of a crystal and its element partitioning behaviour can be used to predict partition coefficients. Depending on the amount of ancillary data available, several approaches can be adopted. First, the data in Table 1 can be used to predict D_i for any 1^+ , 2^+ or 3^+ cation, given a value for the partition coefficient, D_a , of an isovalent cation a (radius r_a) at the pressure, temperature and composition of interest, via the relationship:

$$D_i(P, T, X) = D_a(P, T, X) \times \exp \left[\frac{-4\pi EN_A \left[\frac{r_0}{2} (r_a^2 - r_i^2) + \frac{1}{3} (r_i^3 - r_a^3) \right]}{RT} \right] \quad (3)$$

For example, for $D_a = D_{Ca} = 1$, $T = 1,150^\circ\text{C}$ and the data in Table 1, we predict $D_{Ra} = 0.015$ for anorthite, 0.57 (albite), 0.22 (An_{30}) and 8×10^{-7} (diopside). These values are much smaller than the corresponding values of D_{Ba} , commonly used as a proxy for Ra (ref. 24).

In the absence of partitioning data for an isovalent cation a^{n+} , the partition coefficient for a heterovalent cation b^{m+} ($n \neq m$), can be used for prediction provided that the relationship between $D_0^{n+}(P, T, X)$ and $D_0^{m+}(P, T, X)$ is known. For example, for each clinopyroxene in Table 1, D_0^{3+}/D_0^{2+} is 0.14 ± 0.06 . Using this value together with elastic data from Table 1 and $r_0 = 1.02 \text{ \AA}$ (appropriate for sub-calcic pyroxenes), we have derived a relationship between D_{Ca} and D_{REE} that is in good agreement with the parameterization of Jones²⁵ (Fig. 4).

Finally, for mineral phases where $\bar{E}$ is unavailable, the relationship of ref. 23 (Fig. 3) can be used to estimate $\bar{K}$ (and hence $\bar{E}$) for each valence, simply by assuming that $r_0 = r_y$, where y is the host cation on the site of interest. This information can be combined with D_a , for homovalent reference element a , to predict D_i values using equation (3). As Young's Moduli are relatively insensitive to P - T changes within the Earth²⁶, this approach can be used to predict partition coefficients for high-pressure mantle phases. For example, for MgSiO_3 perovskite at

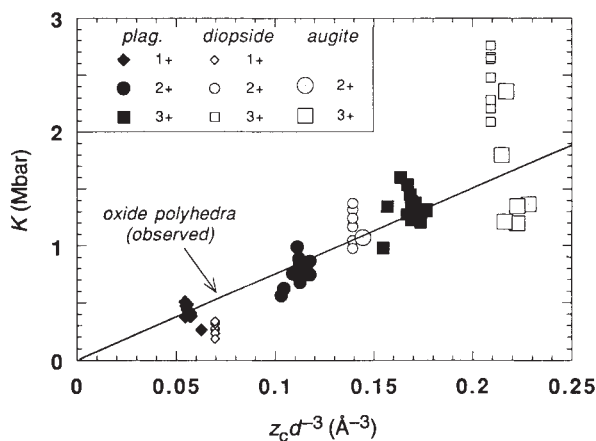


FIG. 3 Comparison of bulk moduli estimated from partitioning experiments (Table 1) with those observed for cation-oxygen polyhedra²³. Values of $\bar{E}$ for diopside and plagioclase were taken from the new experimental data; those for augite were taken from published experimental data^{18,33,34}. In all cases, $\bar{E}$ was converted to $\bar{K}$ using a Poisson ratio of 0.25. Abscissa is z_c/d^3 , where z_c is the mean cation charge and d is the mean cation-oxygen separation calculated as $r_0 + 1.38 \text{ \AA}$, where $1.38 \text{ \AA}$ is the radius of O^{2-} in four-fold coordination³¹. The solid line is the best fit to the cation-oxygen polyhedra data, taken from ref. 23.

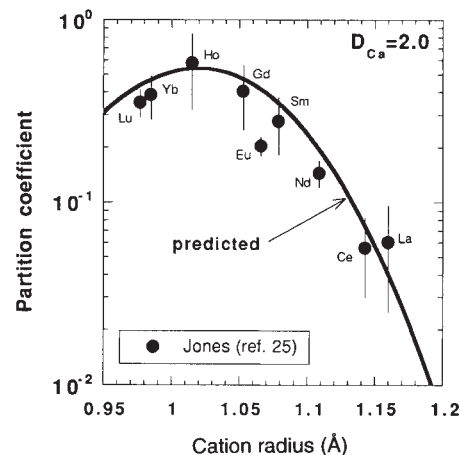


FIG. 4 Comparison of predicted and observed pyroxene partition coefficients for the REE. Data are taken from a parametrization of over 760 experimental data for orthopyroxene, pigeonite and augite²⁵, with D_{Ca} set to 2.0. Error bars are 1 s.d. The solid curve shows the predicted variation, calculated for $D_{Ca} = 2.0$, $r_0 = 1.02 \text{ \AA}$, $D_0^{3+}/D_0^{2+} = 0.14$, $\bar{E}_{\text{cp}}^{3+} = 3,580 \text{ kbar}$ and $\bar{E}_{\text{cp}}^{2+} = 1,820 \text{ kbar}$ (see Table 1).

$2,400^\circ\text{C}$, with $r_0 = r_{\text{MgVIII}} = 0.89 \text{ \AA}$, we predict $D_{\text{Sm}} = 0.19 D_{\text{Sc}}$ and $D_{Ca} = 0.18 D_{\text{Mg}}$, which are in good agreement with experimental data²⁷⁻²⁹.

Exceptions to these simple predictive approaches may include ions with lone pairs of electrons (such as Pb^{2+} , Tl^{+}) or polarizable d shells (Hg^{+} , Cd^{2+} , for example), and transition ions with appreciable crystal field stabilization energies³⁰, like Ni^{2+} . □

Received 17 May; accepted 31 October 1994.

- Irving, A. J. *Geochim. cosmochim. Acta* **42**, 743-770 (1978).
- Henderson, P. *Inorganic Geochemistry* (Pergamon, Oxford, 1982).
- Beattie, P. D. et al. *Geochim. cosmochim. Acta* **57**, 1605-1606 (1993).
- Takahashi, E. & Irvine, T. N. *Geochim. cosmochim. Acta* **45**, 1181-1185 (1981).
- Jones, J. H. & Burnett, D. S. *Geochim. cosmochim. Acta* **51**, 769-782 (1987).
- Nagasawa, H. *Science* **152**, 767-769 (1966).
- Tsang, T., Philpotts, J. A. & Yin, L. J. *Phys. Chem. Solids* **39**, 439-442 (1978).
- Brice, J. C. J. *Crystal Growth* **28**, 249-253 (1975).
- Blundy, J. D. & Wood, B. J. *Geochim. cosmochim. Acta* **55**, 193-209 (1991).
- Blundy, J. D. & Wood, B. J. *Mineralog. Mag.* **58A**, 101-102 (1994).
- Beattie, P. *Mineralog. Mag.* **58A**, 63-64 (1994).
- Onuma, N., Higuchi, H., Wakita, H. & Nagasawa, H. *Earth planet. Sci. Lett.* **5**, 47-51 (1968).
- Jensen, B. B. *Geochim. cosmochim. Acta* **37**, 2227-2242 (1973).
- Mackrodt, W. C. in *Computer Simulation of Solids* (eds Catlow, C. R. A. & Mackrodt, W. C.), *Lecture Notes in Physics* **166**, 175-194 (Springer, Berlin, 1982).
- Catlow, C. R. A., James, D., Mackrodt, W. C. & Stewart, R. F. *Phys. Rev.* **B25**, 1006-1026 (1982).
- Smith, J. V. & Brown, W. L. *Feldspar Minerals 1. Crystal Structures, Physical, Chemical and Microtextural Properties* (Springer, Berlin, 1988).
- Smyth, J. R. & Bish, D. L. *Crystal Structures and Cation Sites of the Rock Forming Minerals* (Allen & Unwin, Boston, 1988).
- Hart, S. R. & Dunn, T. *Contr. Miner. Petrol.* **113**, 1-8 (1993).
- Liu, C. Q., Masuda, A., Shimizu, H., Takahashi, K. & Xie, G. H. *Geochim. cosmochim. Acta* **56**, 1523-1530 (1992).
- Levien, L. & Prewitt, C. T. *Am. Miner.* **66**, 315-323 (1981).
- Angel, R. J., Hazen, R. M., McCormick, T. C., Prewitt, C. T. & Smyth, J. R. *Phys. Chem. Miner.* **15**, 313-318 (1988).
- Anderson, D. L. & Anderson, O. L. *J. geophys. Res.* **75**, 3494-3500 (1970).
- Hazen, R. M. & Finger, L. W. *J. geophys. Res.* **84**, 6723-6728 (1979).
- Rubin, K. H. & Macdougall, J. D. *Nature* **335**, 158-161 (1988).
- Jones, J. H. in *Handbook of Geophysical Constants* (ed. Ahrens, T. J.) Vol. 3, Ch. 7 (American Geophysical Union, Washington, in the press).
- Sumino, Y. & Anderson, O. L. in *Handbook of Physical Properties of Rocks* (ed. Carmichael, R. S.) Vol. 3, 39-138 (Chemical Rubber Company, Boca Raton, 1989).
- Kato, T., Ringwood, A. E. & Irfune, T. *Earth planet. Sci. Lett.* **90**, 65-68 (1988).
- Drake, M. J., McFarlane, E. A., Gasparik, T. & Rubie, D. C. *J. geophys. Res.* **98**, 5427-5431 (1993).
- Agee, C. B. *Nature* **346**, 834-837 (1990).
- Burns, R. G. *Mineralogical Applications of Crystal Field Theory* (Cambridge University Press, Cambridge, 1970).
- Shannon, R. D. *Acta crystallogr.* **A32**, 751-767 (1976).
- Hinton, R. W. *Chem. Geol.* **83**, 11-25 (1990).
- McKay, G., Wagstaff, J. & Tang, S. R. *Geochim. cosmochim. Acta* **50**, 927-937 (1986).
- Grutzeck, M., Kriedelbaugh, S. & Weill, D. F. *Geophys. Res. Lett.* **1**, 273-275 (1974).

ACKNOWLEDGEMENTS. J.D.B. acknowledges receipt of a NERC Fellowship. We thank E. Wasserman, N. Allan, K. Fawcett and G. Helffrich for discussion. R. Angel for critically reading the manuscript, J. Jones for unpublished data, and J. Craven and R. Hinton (of the NERC Scientific Services ion-probe facility at Edinburgh University) for analytical assistance.

Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria

Petra Rueter*, Ralf Rabus*, Heinz Wilkes†, Frank Aeckersberg*, Fred A. Rainey‡, Holger W. Jannasch§ & Friedrich Widdel*||

* Max-Planck-Institut für Marine Mikrobiologie, Fahrenheitstr. 1, D-28359 Bremen, Germany

† Institut für Erdöl und Organische Geochemie (ICG-4), Forschungszentrum Jülich GmbH, Leo-Brandt-Strasse, D-52428 Jülich, Germany

‡ Deutsche Sammlung von Mikroorganismen, Mascheroder Weg 1b, D-38124 Braunschweig, Germany

§ Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

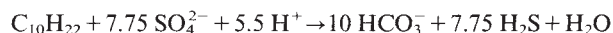
|| To whom correspondence should be addressed

MANY crude oil constituents are biodegradable in the presence of oxygen; however, a substantial anaerobic degradation has never been demonstrated^{1,2}. An unusually low content of *n*-alkanes in oils of certain deposits is commonly attributed to selective utilization of these hydrocarbons by aerobic microorganisms^{3,4}. On the other hand, oil wells and production fluids were shown to harbour anaerobic sulphate-reducing bacteria⁵⁻⁸, but their actual electron donors and carbon sources were unknown. On the basis of nutritional properties of various bacterial isolates it was assumed that fatty acids and H₂ are potential electron donors for sulphate reduction *in situ*⁵⁻⁸. Here we demonstrate that hydrocarbons in crude oil are used directly by sulphate-reducing bacteria growing under strictly anoxic conditions. A moderately thermophilic pure culture selectively utilizes *n*-alkanes in oil for sulphate reduction to sulphide. In addition, a mesophilic sulphate-reducing enrichment culture is shown to oxidize alkylbenzenes in oil. Thus, sulphate-reducing bacteria utilizing aliphatic and aromatic hydrocarbons as electron donors may present a significant source of sulphide in oil deposits and oil production plants.

Enrichment of hydrocarbon-oxidizing sulphate-reducing bacteria on crude oil was attempted in strictly anoxic artificial medium with the same concentration of sulphate (28 mM), chloride, sodium, potassium, magnesium and calcium ions as

natural sea water⁹. Inocula were from Guaymas Basin sediment and the water phase of a North Sea oil tank at Wilhelmshaven. Guaymas Basin is part of the tectonic spreading zone in the Gulf of California where hydrothermal activity leads to the formation of aliphatic and aromatic hydrocarbons from diatomaceous detritus^{10,11}. The sediment samples were taken at a depth of 2,000 m by means of the research submersible *Alvin* during Cruise 125, Leg 22, of the Woods Hole Oceanographic Institution. Turnover studies of sulphate in these sediments indicated high *in situ* activities of sulphate-reducing bacteria around 30, 60, 90 and 105 °C (ref. 12). Thus far, aerobic but no anaerobic hydrocarbon-degrading bacteria could be isolated from Guaymas Basin sediment¹³.

Anoxic enrichments with Guaymas Basin sediment and crude oil were most effective at 60 °C, producing more than 10 mM sulphide within 2–3 weeks. Further cultivation was successful on *n*-decane. A pure culture, strain TD3, was isolated from colonies in anoxic agar dilution series⁹ containing caproate (2 mM) as a water-soluble electron donor; the isolate was then grown again on decane. Strain TD3 consisted of slowly motile, often slightly curved rods (2 × 0.8 µm). Growth was optimal between 55 and 65 °C and at a pH of around 6.8. Desulfovibrin, a pigment common to *Desulfovibrio* species, was not detectable in strain TD3 using a fluorescence test¹⁴. Phylogenetically, strain TD3 exhibited a new, deep branch within the sulphate-reducing eubacteria of the delta subdivision (Fig. 1). Strain TD3 grew on crude oil samples from three different locations (three land-based oil fields, in the Hamburg–Hannover region). In growth tests with defined compounds, strain TD3 oxidized *n*-alkanes from C₆ to C₁₆, with best growth in the C₈ to C₁₂ range. Fatty acids from C₄ to C₁₈ were also utilized, but no H₂, ethanol or lactate, the common substrates of *Desulfovibrio* and many other sulphate-reducing bacteria⁹. Alkane oxidation by strain TD3 was quantified in anoxic vials^{9,15} with *n*-decane (12% dissolved in pristane as the inert carrier phase and internal standard for gas chromatography). For every 1.5 mmol of decane consumed per litre of culture medium, 11.3 mM H₂S was formed, in agreement with the reaction



Sulphide production by strain TD3 with crude oil as the sole organic substrate is shown in Fig. 2a. Sulphide production and growth never occurred without either oil or sulphate. Growth was accompanied by emulsification of the oil phase, probably because of the pronounced attachment of cells. The concentration of produced sulphide increased with the amount of oil added

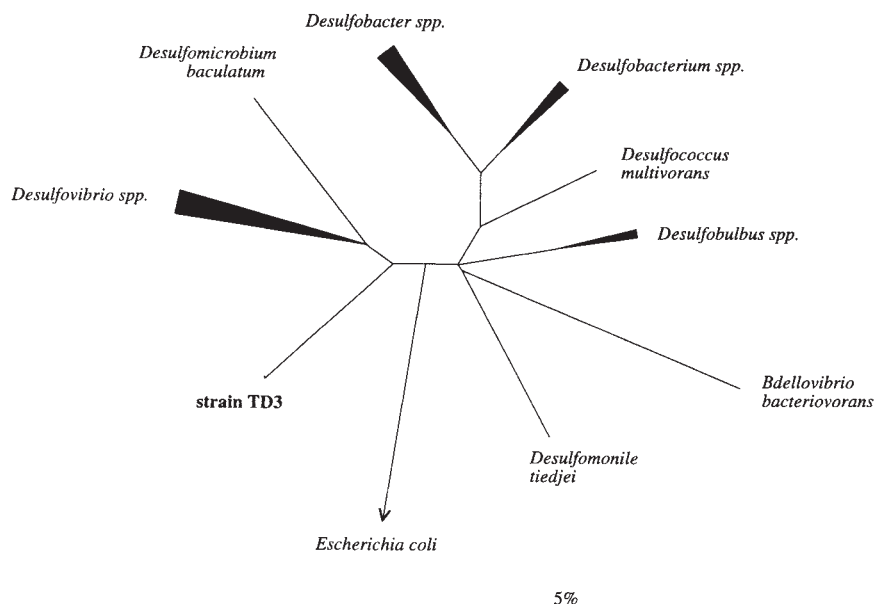
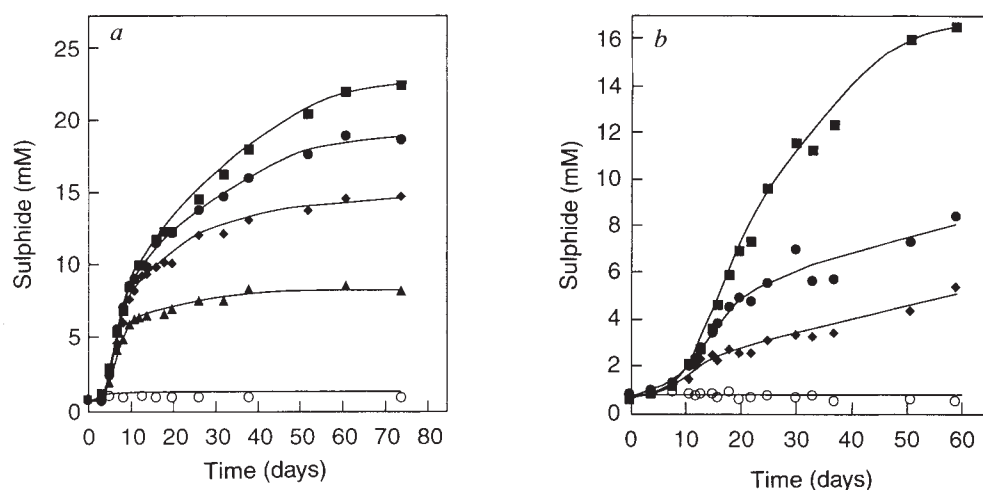


FIG. 1 Reconstructed phylogenetic tree showing the position of strain TD3 within the delta subclass of the Proteobacteria. Amplification of the 16S rRNA gene and sequencing of double-stranded templates has been described^{22,23}. The almost complete 16S rDNA gene sequence (1,441 nucleotides) of strain TD3 (accession number EMBL X80922) was compared with published bacterial 16S rRNA sequences. The sequence was aligned manually and the alignment optimized on the basis of secondary structure. Corrected pairwise evolutionary distances were computed and the least squares distance method employed in the reconstruction of a phylogenetic tree^{24,25}. The 16S rRNA sequence of *Escherichia coli* (gamma subclass of the Proteobacteria) was used to establish the root of this tree.

FIG. 2 Sulphide production by sulphate-reducing bacteria growing with various amounts of crude oil as the only organic substrate. *a*, Thermophilic pure culture (strain TD3) utilizing alkanes from crude oil. Amounts of oil (from oil production plant near Hamburg) per 115 ml of medium: open circles, no oil; triangles, 0.2 ml; diamonds, 0.5 ml; filled circles, 1.0 ml; squares, 3.0 ml. *b*, Mesophilic enrichment culture utilizing aromatic hydrocarbons from crude oil. Amounts of oil (from North Sea oil tank, Wilhelmshaven) per 115 ml of medium: open circles, no oil; diamonds, 0.5 ml; filled circles, 1.0 ml; squares, 4.0 ml. The sulphide produced did not increase proportionally with the amount of oil added; the reason for the lack of linear correlation is as yet unknown. Experiments were done in bicarbonate-buffered (pH 7) artificial sea water medium, prerduced with sodium sulphide (1 mM) and dithionite (0.15 mM)⁹. The cultivation vessels were special flat vials (130 ml) with a small butyl-rubber-sealed orifice¹⁵. The head space above the medium (115 ml) contained an anoxic nitrogen/carbon dioxide mixture (90/10, vol/vol). Stocks of crude oil (120 ml) were deaerated by repeated shaking under nitrogen in bottles (250 ml) provided with a side arm for aseptic gas supply and



an orifice sealed with Teflon-coated rubber^{9,15}. Oil was then autoclaved and stored in the same tightly sealed bottles under nitrogen. Gas chromatography did not reveal any changes of hydrocarbon patterns due to this procedure. Anoxic oil was aseptically injected into the incubation vials through nitrogen-flushed syringes. The orifice of all cultivation vials was always kept in a downward position below the medium surface to avoid contact of the oil with the stopper^{9,15,16}. Sulphide was analysed in a colorimetric microassay¹⁵.

per culture volume. With 3.0 ml oil per 115 ml medium, more than 20 mM sulphide was produced within 74 days. The rate of sulphide production was highest (10 mM in 11 days) during the initial growth phase. If the average C/H ratio of utilized hydrocarbons is approximated by the formula of decane, the H₂S concentration (7.5 mM) formed by sulphate reduction with a small amount of crude oil (density 0.884 kg l⁻¹; 0.2 ml added per 115 ml medium) leads to an estimate of 9% (w/w) which is oxidizable by strain TD3. Comparative gas chromatographic analyses of oil from growth cultures and control experiments without sulphate revealed that *n*-alkanes from C₈ to C₁₁ disappeared during sulphate reduction (Fig. 3). Higher alkanes up to a chain length of C₁₆ were partially consumed. The consumption pattern matches the total as well as the preferred range of alkanes utilized in growth tests with defined compounds. Strain TD3 did not alter the pattern of aromatic hydrocarbons (not shown). The alkane composition of the oil from the sulphate-free control was exactly the same as in the original oil (not shown), indicating that volatilization of hydrocarbons during incubation can be excluded. This is also confirmed by the observation that neither undegraded volatile branched alkanes nor volatile aromatic hydrocarbons were lost in cultures of strain TD3.

The growth experiments with strain TD3 confirm the capacity of specialized sulphate-reducing bacteria for strictly anaerobic oxidation of alkanes to carbon dioxide, as demonstrated before with strain Hxd3, a mesophilic isolate¹⁵. As alkane utilization (C₁₂ to C₂₀) by the latter had previously been shown only with pure compounds, this strain was also tested with crude oil. In contrast to strain TD3, strain Hxd3 exhibited only poor growth with crude oil as the only substrate, yielding 4 mM sulphide in 2 months. Described species of sulphate-reducing bacteria⁹ that utilize fatty acids grew neither with pure alkanes nor with crude oil.

With inocula from the North Sea oil-tank, only mesophilic enrichments (30 °C) developed with crude oil as substrate. The initially clear medium became turbid, because of bacterial growth, without obvious emulsification of the oil. Various oval to rod shaped bacteria were observed, which partly occurred in aggregates. If controls without sulphate or oil were inoculated, neither growth nor sulphide production was observed. Produced

sulphide concentrations increased with the amount of oil added (Fig. 2b). Sulphate reduction was associated with an alteration of the aromatic hydrocarbon fraction (Fig. 4). The alkane fraction remained unchanged (not shown). Toluene and *o*-xylene disappeared completely; *m*-xylene, *o*-ethyltoluene and *m*-ethyltoluene were partially consumed. There was no significant consumption of other aromatic hydrocarbons. Sulphate-reducing subcultures were obtained from the oil-enrichment on toluene, *o*-xylene and *m*-xylene under cultivation conditions described recently¹⁶, but not on benzene or *p*-xylene.

Oxidation of toluene and the three isomers of xylene with sulphate as electron acceptor has been demonstrated with defined compounds added to enrichments from aquifers^{17,18}. A pure culture of a sulphate-reducing bacterium, *Desulfobacula toluolica*, oxidized toluene to carbon dioxide¹⁶. The latter was tested on crude oil, but only about 1 mM sulphide was produced in 2 months.

The origin of hydrogen sulphide (H₂S) found in petroleum deposits has been attributed to several processes, owing to the variation in sulphur isotope fractionation (δ³⁴S values) in different reservoirs¹⁹. A chemical sulphate reduction with organic oil components as reductants is thought to occur at temperatures above about 120 °C; the reactive intermediate is most probably elemental sulphur from oxidation of sulphide with sulphate²⁰. Another chemical source may be a thermal cleavage of organic sulphur compounds in oil²⁰. Chemical formation of sulphide is associated only with small isotopic fractionation¹⁹. High fractionation, which is indicative of bacterial sulphate reduction²¹, has been measured in reservoirs with temperatures not higher than about 60 °C (ref. 19), the temperature optimum of strain TD3. The recent finding of extremely thermophilic sulphate reducers in production fluids⁸ and sulphate reduction in hydrothermal sediments¹² demonstrate that biological sulphate reduction in oil reservoirs is possible up to temperatures of around 100 °C. Biological sulphide production with crude oil as the only substrate has been observed in a mesophilic¹⁵ and an extremely thermophilic⁸ enrichment. But so far the actual compounds in oil leading to sulphide production in such enrichments and in anoxic oil reservoirs at any temperature range were unknown. Nutritional properties of various sulphate-reducing bacteria isolated from oil fields suggested that petro-

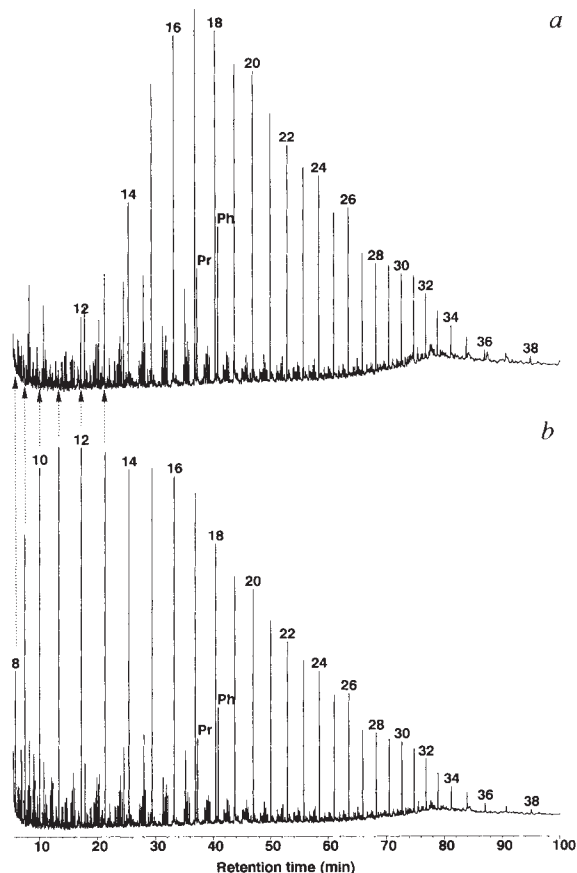


FIG. 3 Gas chromatograms of the saturated hydrocarbon fraction of crude oil after anaerobic incubation with strain TD3. *a*, From culture with sulphate as the electron acceptor; *b*, from control without sulphate (magnesium sulphate replaced by magnesium chloride). The anoxic incubation method was essentially the same as in Fig. 2, except that the culture volume was increased. A volume of 1.2 ml crude oil (from oil field near Hamburg) was incubated in 420 ml medium for 30 days at 60 °C. Numbers on the peaks indicate chain lengths of *n*-alkanes; Pr, pristane; Ph, phytane. Analyses were done on a HP 5890 instrument equipped with an Ultra 1 fused silica capillary column (50 m, internal diameter 0.21 mm, coating 0.33 μ m). The temperature programme was run from 90 °C (4 min isotherm) to 310 °C, at 3 °C min⁻¹. Fractionation of oil samples before analysis was by medium pressure liquid chromatography²⁶.

leum fatty acids^{6,7}, hydrogen from chemical and bacterial processes^{6,8}, or polar products from aerobic hydrocarbon-degrading bacteria⁵ are potential electron donors for sulphate reduction to sulphide *in situ*. The possibility of sulphate reduction with hydrocarbons from oil has been speculated about^{3,15}, but this is the first experimental demonstration that sulphate-reducing bacteria can indeed utilize aliphatic and aromatic hydrocarbons directly from oil under anoxic conditions. Sulphate-reducing bacteria with the capacity for hydrocarbon oxidation may have contributed to sulphide production and specific hydrocarbon depletion during early maturation of various oil reservoirs. Moreover, such bacteria may also be involved in specific hydrocarbon degradation in anoxic petroleum-bearing

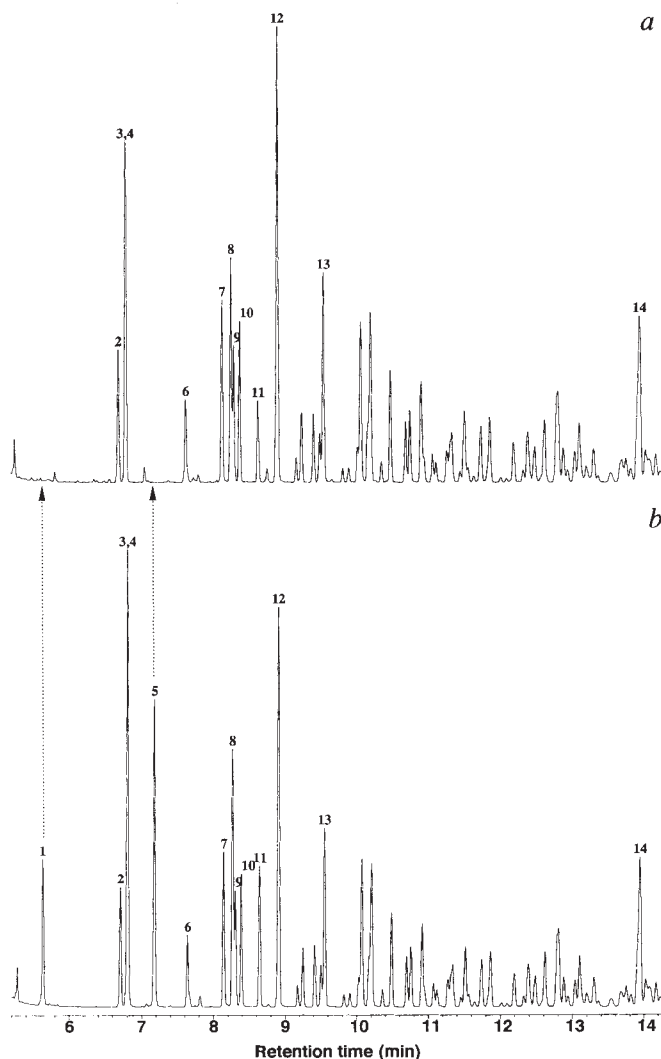


FIG. 4 Partial gas chromatograms of the aromatic hydrocarbon fraction of crude oil after anaerobic incubation with cells from a sulphate-reducing enrichment culture. *a*, From culture with sulphate as electron acceptor; *b*, from control without sulphate. 1, Toluene; 2, ethylbenzene; 3, *m*-xylene; 4, *p*-xylene; 5, *o*-xylene; 6, *i*-propylbenzene; 7, propylbenzene; 8, *m*-ethyltoluene; 9, *p*-ethyltoluene; 10, 1,3,5-trimethylbenzene; 11, *o*-ethyltoluene; 12, 1,2,4-trimethylbenzene; 13, 1,2,3-trimethylbenzene; 14, naphthalene. The analysed oil (from North Sea oil tank) was from the experiment shown in Fig. 2b, upper curve, and a sulphate-free control. Analyses were done on a HP 5890 instrument equipped with an Ultra 2 fused silica capillary column (50 m, internal diameter 0.21 mm, coating 0.33 μ m). The temperature programme was run from 90 °C (4 min isotherm) to 120 °C, at 50 °C min⁻¹ and further from 120 °C (1 min isotherm) to 310 °C, at 3 °C min⁻¹. Structural assignment of the indicated compounds is based on comparison of their gas chromatographic and mass spectrometric behaviour with that of authentic standards. Separation of *m*-xylene and *p*-xylene was achieved by headspace gas chromatography under isothermic conditions (not shown). Fractionation of oil samples before analyses was by medium pressure liquid chromatography²⁶.

marine sediments, and in the souring of oil wells upon flooding with sulphate-rich sea water. □

Received 4 August; accepted 25 October 1994.

1. Atlas, R. M. *Microbiol. Rev.* **45**, 180–209 (1981).
2. Leahy, J. & Colwell, R. R. *Microbiol. Rev.* **54**, 305–315 (1990).
3. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence* 2nd edn (Springer, Berlin, New York, 1984).
4. Blanc, P. & Connan, J. in *Applied Petroleum Geochemistry* (ed. Bordenave, M. L.) 151–174 (Editions Technip, Paris, 1993).

5. Nazina, T. N., Rozanova, E. P. & Kuznetsov, S. I. *Geomicrobiol. J.* **4**, 103–130 (1985).
6. Cord-Ruwisch, R., Kleinitz, W. & Widdel, F. *J. Petrol. Technol.* 97–106 (January 1987).
7. Rosnes, J. T., Torsvik, T. & Lien, T. *Appl. Environ. Microbiol.* **57**, 2302–2307 (1991).
8. Stetter, K. O. *et al. Nature* **365**, 743–745 (1993).
9. Widdel, F. & Bak, F. in *The Prokaryotes* 2nd edn Vol. 4 (eds Balows, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H.) 3352–3378 (Springer, Berlin, New York, 1992).
10. Simoneit, B. R. T. & Lonsdale, P. F. *Nature* **295**, 198–202 (1982).

11. Bazyliński, D. A., Farrington, J. W. & Jannasch, H. W. *Org. Geochem.* **12**, 547–558 (1988).
12. Jørgensen, B. B., Isaksen, M. F. & Jannasch, H. W. *Science* **258**, 1756–1757 (1992).
13. Bazyliński, D. A., Wirsén, C. O. & Jannasch, H. W. *Appl. environ. Microbiol.* **55**, 2832–2836 (1989).
14. Postgate, J. R. *The Sulphate-reducing Bacteria* 2nd edn (Cambridge University Press, Cambridge, 1984).
15. Aeckersberg, F., Bak, F. & Widdel, F. *Arch. Microbiol.* **156**, 5–14 (1991).
16. Rabus, R., Nordhaus, R. & Widdel, F. *Appl. environ. Microbiol.* **59**, 1444–1451 (1993).
17. Beller, H. R., Grbić-Galić, D. & Reinhard, M. *Appl. environ. Microbiol.* **58**, 786–793 (1992).
18. Edwards, E. A., Wills, L. E., Reinhard, M. & Grbić-Galić, D. *Appl. environ. Microbiol.* **58**, 794–800 (1992).
19. Nielsen, H. et al. in *SCOPE 43, Stable Isotopes* (eds Krouse, H. R. & Grinenko, V. A.) 65–132 (Wiley, New York, 1991).
20. Orr, W. *Amer. Assoc. Petrol. Geologists Bull.* **58**, 2295–2318 (1974).

21. Schidlowski, M., Hayes, J. M. & Kaplan, I. R. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 149–186 (Princeton University Press, Princeton, 1983).
22. Rainey, F. A., Dorsch, M., Morgan, H. W. & Stackebrandt, E. *Syst. appl. Microbiol.* **15**, 197–202 (1992).
23. Rainey, F. A. & Stackebrandt, E. *FEMS Microbiol. Lett.* **113**, 125–128 (1993).
24. Jukes, T. H. & Cantor, C. R. *Mammalian Protein Metabolism* (Academic, New York, 1969).
25. De Soete, G. *Psychometrika* **48**, 621–626 (1983).
26. Radke, M., Willsch, H. & Welte, D. H. *Analyt. Chem.* **52**, 406–411 (1980).

ACKNOWLEDGEMENTS. This research was supported by the Max-Planck-Gesellschaft, Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie, US National Science Foundation and Office of Naval Research. We thank W. Kleinitz (Lingen) and J. Fischer (Wilhelmshaven) for oil samples, B. Horsfield and R. G. Schaefer for discussion, and U. Disko and H. Willsch for technical assistance.

Evolution of homeotic gene regulation and function in flies and butterflies

Robert W. Warren, Lisa Nagy, Jane Selegue, Julie Gates & Sean Carroll*

Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA

It has been proposed that the evolution of homeotic genes parallels, and to some degree directs, the evolution of segment diversity in the myriapod-insect lineage^{1–3}. But the discovery of discrete Antennapedia complex (ANT-C) and bithorax complex (BX-C) gene members in crustacea⁴, chelicerates⁵, annelids^{6–8} and various insects^{9–11}, as well as in vertebrates¹², indicates that the expansion and diversification of homeotic genes preceded the diversification of arthropods and insects. How, then, have these genes influenced the evolution of body plans? To address this question, we now examine homeotic gene expression and regulation in butterflies (Lepidoptera), which, unlike flies, possess larval abdominal limbs and two pairs of wings. We show that the difference in larval limb number between these insects results from striking changes in BX-C gene regulation in the butterfly abdomen, and we deduce that the wing-patterning genes regulated by *Ultrabithorax* have diverged in the course of butterfly and fly evolution. These findings have general implications for the role of homeotic genes in animal evolution.

To examine homeotic gene expression in butterflies, *Precis coenia Antennapedia* (*Antp*), *Sex combs reduced* (*Scr*), *abdominal-A* (*abd-A*) and *Ultrabithorax* (*Ubx*) complementary DNAs were isolated from a *P. coenia* embryonic cDNA library. Alignment of the deduced amino-acid sequences from the *P. coenia* cDNAs with the *Drosophila* homeotic protein sequences established the identity of each gene (Fig. 1a). As the differences in limb and wing number between flies and butterflies arise in the segments in which *Ubx* and *abd-A* are expressed, we first asked whether the expression of these BX-C genes differed in *P. coenia* and *Drosophila*. When embryogenesis is ~20% completed and all of the trunk segments of the *P. coenia* embryo are well formed, the *P. coenia Ubx* (Fig. 1b, c; Fig. 2a), and *abd-A* (Fig. 1d, e) genes appear to be expressed in anteroposterior domains and parasegmental registers as in *Drosophila*^{13–15}, although the *Ubx* gene seems to be expressed at very low levels within posterior T2 and T3 (Fig. 2a). The conservation of the BX-C and ANT-C (not shown) homeotic gene expression domains indicates that the embryonic body plans of butterflies and fruitflies are fundamentally similar and cannot explain the observed differences in thoracic and abdominal appendage number.

At ~20% of *P. coenia* embryogenesis, the larval thoracic limbs are forming and express the *Distal-less* (*Dll*) limb-patterning gene at high levels¹⁶; there is no sign yet of the abdominal prolegs

(Fig. 2b). At this stage, *abd-A* protein (Fig. 2a) and RNA (Fig. 1d) are expressed throughout the more anterior abdominal segments (including A3–A6, where larval prolegs will eventually develop) and high levels of *Antp* expression are restricted to the thorax (not shown). After 20% of embryogenesis, however, the patterns of *abd-A*, *Dll* and *Antp* expression change dramatically (Fig. 2c–h). The first sign of the abdominal proleg anlage is the repression of *abd-A* protein (Fig. 2c) and RNA (not shown) expression in a ventrolateral patch of cells in abdominal segments A3–A6 (Fig. 2c). At this stage, neither *Dll* (Fig. 2d) nor *Antp* (Fig. 2e) are expressed in the abdominal proleg anlage. As each patch of *abd-A* repression enlarges in A3–A6 (Fig. 2f), *Dll* (Fig. 2g) and *Antp* (Fig. 2h) expression are both activated in the abdominal limb primordia.

To determine whether *Dll* could be responsible for, or could follow, the downregulation of *Ubx/abd-A*, double-label experiments using antibodies against *P. coenia Dll* and the *Ubx/abd-A* antibody were performed. The activation of *Dll* in the abdomen within the proleg anlage is clearly trailing the repression of *Ubx* and *abd-A* expression (Fig. 3a, b), indicating that BX-C gene repression is the initial event. As *Ubx* and *abd-A* repress *Dll* expression in the abdominal segments of the *Drosophila* embryo¹⁷, it appears that abdominal limb formation in butterflies has been made possible by the evolution of a regulatory mechanism for shutting off these two BX-C genes in selected cell populations, which then permits *Dll*, and *Antp*, to be expressed.

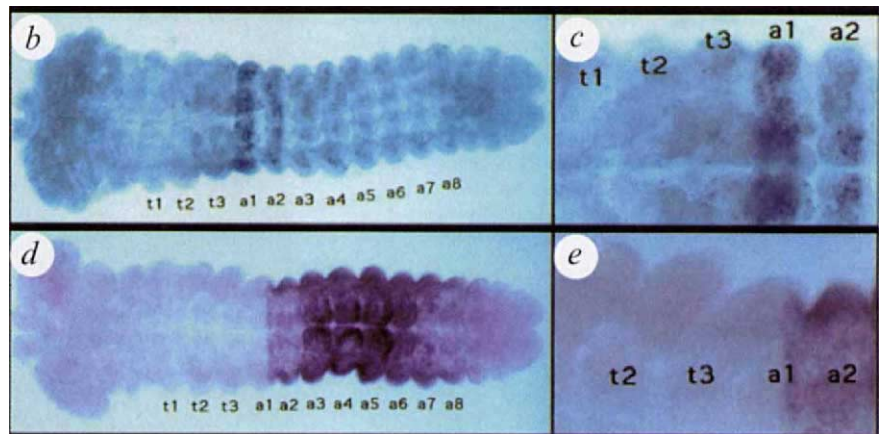
The differences in abdominal limb formation in flies and butterflies could have been due to differences in the number, regulation or function of homeotic genes. Evolution at the level of gene number was unlikely in light of studies that identified BX-C genes in other insects^{9–11} and arthropods^{4,5}, and is excluded by our data. On the other hand, the expression of the BX-C target gene, *Dll*, in the abdomen could be accomplished by altering the regulation of *Dll* in the abdomen (for example by mutations in *Dll* cis-regulatory elements operating in A3–A6). So why have butterflies repressed BX-C genes in selected cell populations within a group of abdominal segments? The most logical answer is that the derepression of *Dll* is probably not sufficient to elaborate a proleg (or any other limb for that matter). That is, an entire programme of genes involved in limb formation must be derepressed in order to form a proleg. To release the limb genetic programme, the BX-C homeotic genes must be repressed.

The phylogenetic distribution, anatomy and development of prolegs suggests that these must be very basic or phylogenetically old structures, distinct from imaginal legs, and derived perhaps from an ancestor common to insects and myriapods¹⁸. Birket-Smith¹⁸ has even proposed, on largely anatomical grounds, such as muscle organization, that prolegs could be related to the lobopods of primitive Onychophora. As prolegs are found in several insect orders, it is possible, if not likely, that the selective repression of BX-C gene expression is not just a Lepidopteran invention. But these phenomena in Lepidoptera cannot be extrapolated to the relationship between BX-C genes and limb formation in insect ancestors. There are several plausible ways in which BX-C genes could operate in animals such as myriapods which possess limbs on every trunk segment. It may be that BX-C genes do not regulate the limb genetic programme in myria-

* To whom correspondence should be addressed

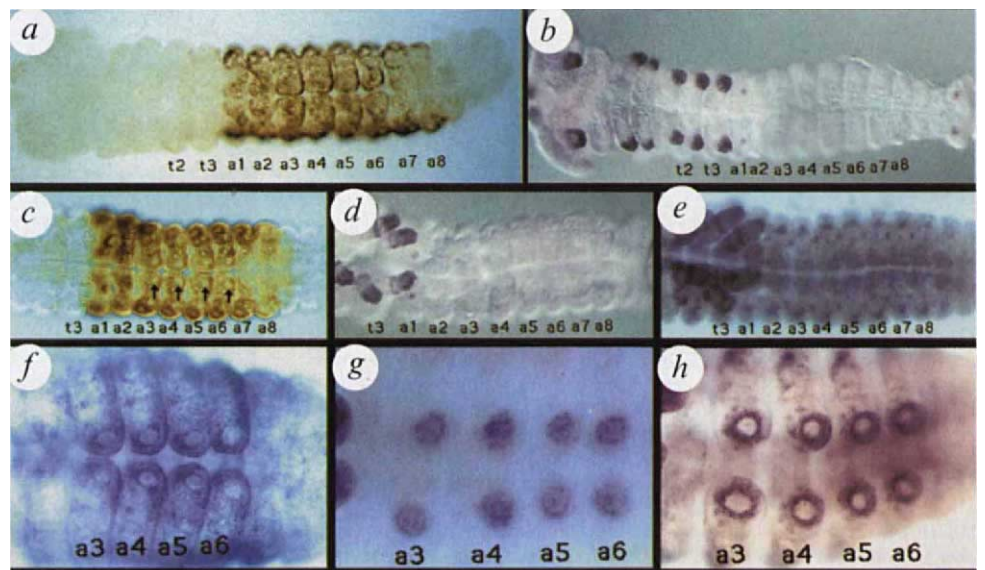
FIG. 1 Structure and expression of *P. coenia* homeotic genes. *a*, The deduced amino-acid sequences obtained from *P. coenia* cDNAs encoding the *Antp*, *Scr*, *Ubx* and *abd-A* genes are aligned with the *Drosophila melanogaster* *Antp* protein. The region overlined corresponds to the homeodomain. Sequence identity with *Antp* is represented as a dot, changes by a letter (upper case for a change that is conserved between respective *Drosophila* and *P. coenia* proteins, lower case for a unique *P. coenia* residue). The *Scr*, *abd-A* and *Ubx* sequences are each truncated in the homeodomain but can be identified on the basis of both homeodomain and C-terminal sequences (note the extensive conservation of C-terminal *P. coenia* *Ubx* residues). *b–e*, The early patterns of *P. coenia* *Ubx* (*b, c*) and *abd-A* (*d, e*) gene expression at about 20% of embryogenesis are shown at low (*b, d*) and high magnification (*c, e*). *Ubx* appears to be expressed at the highest level in anterior A1 with significant but diminishing levels in more posterior segments. While not pronounced at the RNA level, we do detect some expression of *Ubx* protein (for example, see Fig. 2*a*) in lateral regions of T2 and T3. This is similar but not identical to the level and pattern of *Ubx* expression in *Drosophila*. The anterior limit of *abd-A* expression is in posterior A1 and the overall pattern of *abd-A* expression is quite similar to that found in *Drosophila*. The T3 leg overlies part of the A1 segment and is not expressing *Ubx* or *abd-A*.

METHODS. The *P. coenia* *Scr*, *Antp* and *abd-A* homologues were obtained by low stringency hybridization to a *P. coenia* embryonic cDNA library using a ³²P-labelled, 240-bp *Drosophila* *Ubx/abd-A* homeodomain probe (gift from A. Laughon). We obtained six *Scr* clones, seven *Antp* clones, and two *abd-A* clones. Analysis of multiple cDNAs for these genes indicated that each is represented by a single locus. The *P. coenia* *Ubx* homologue was obtained using a *Manduca sexta* *Ubx* probe (159-bp probe encoding the last 8 amino acids of the *Ubx* homeodomain to the C terminus; gift from R. Booker). Low-stringency hybridizations to a *P. coenia* λ gt10 embryonic cDNA library were



performed at 35 °C in 10 ml hybridization solution: 5 × SSC; 0.5% SDS; 5 × Denhardt's; 500 µg ml⁻¹ ssDNA; 10 mM EDTA, pH 8; 50 mM PIPES, pH 7; 43% formamide. Probe concentration was 1 × 10⁶ d.p.m. ml⁻¹ hybridization solution. *Precis coenia* cDNAs were subcloned into pBluescript KS (Stratagene). Because *Scr*, *Ubx* and *abd-A* clones were truncated in the coding region, sequencing was accomplished by end runs of the universal and reverse primer by the Sanger dideoxynucleotide method²⁹ using T7 DNA polymerase³⁰ (US Biochemicals). Additional primers were required to complete the *abd-A* sequence: 5'-GCCGTCAGAAAGATC and 3'-CGACGTATGAAG. Sequencing of the *Antp* clones required degenerate internal primers: Forward primer (homeobox residues 16–21, ELEKEF) 5'-GGAATTCGAA/GCTTGAA/GAAA/GGAA/GTT-3'; reverse primer (homeobox residues 49–54, WFNRR) 5'-GCTCTAGACGICGA/GTTTGA/GAACC-3'. Specific primers were then used to sequence the remaining coding sequence: 5'-CGCTGATAACGAAG and 3'-GGGTTGTGACGAGG. *In situ* hybridization is described in ref. 16.

FIG. 2 Repression of BX-C genes precedes the activation of limb-patterning genes and *Antp* in abdominal limb primordia. *a*, Embryo at 20% of embryogenesis stained with a monoclonal antibody to the *Drosophila* *Ubx/abd-A* proteins³¹ that crossreacts with *P. coenia*, protein expression is throughout the anterior abdomen and *Ubx* is present in lateral regions of T2 and T3. *b*, Embryo between 15 and 20% of embryogenesis expresses the *P. coenia* *Dll* gene in the thoracic and cephalic limb primordia and in small clusters of cells on A1 and near the anus, there is no expression in abdominal segments A2–A10. *c–e*, Embryos at approximately 25% of embryogenesis. *c*, *Ubx/abd-A* protein expression is repressed in small patches of cells (arrows) in abdominal segments A3–A6. *d*, *Dll* RNA, and *e*, *Antp* RNA expression is not yet detectable within small patches of cells in A3–A6 along the ventral midline. The abdominal *Antp* expression along the ventral midline is primarily in the CNS. *Antp* expression at 15% of embryogenesis is limited to T1, T2 and T3 (not shown). *f–h*, Embryos at approximately 40% of embryogenesis. *f*, *abd-A* RNA expression is repressed in large circular patches of cells corresponding to the abdominal leg primordia (a3–a6) which also express *Dll* RNA (*g*), and *Antp* RNA (*h*). The plane of *Antp* staining shown is at the level of the body wall, *Antp* expression at this stage is limited to the proximal part of the proleg. It appears that *Antp* expression in the prolegs is at the junction



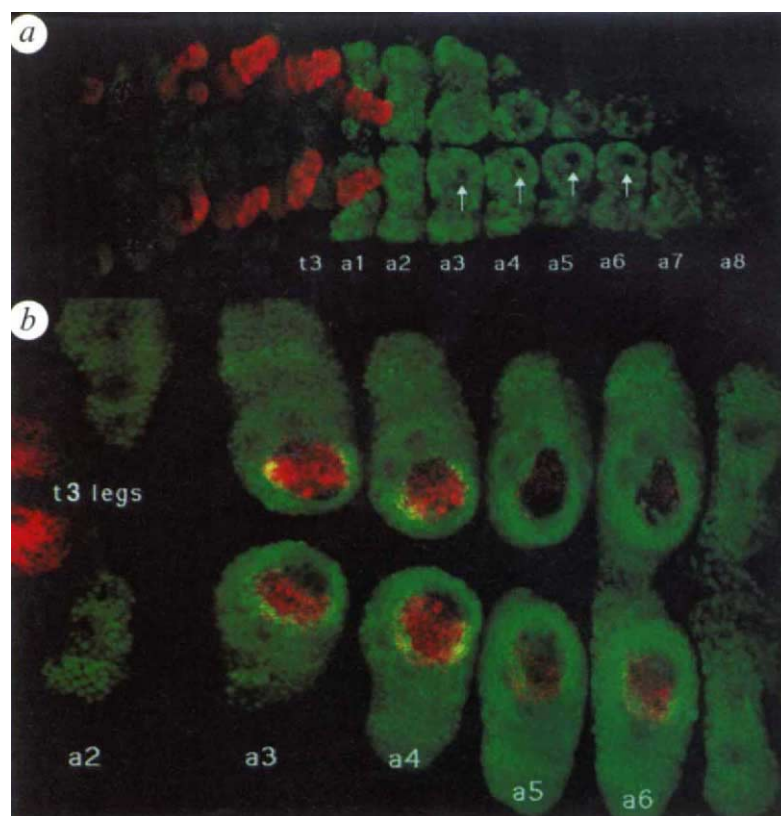
of *abd-A/Ubx*- and *Dll*-expressing cells, and overlaps the edges of both domains. The anal prolegs also express *Dll* and *Antp* (not shown). **METHODS.** *In situ* hybridization (*b, d–h*) was performed as in Fig. 1. Antibody staining with the monoclonal antibody which recognizes the *Ubx* and *abd-A* proteins of flies³¹, locusts³¹, and shown here, butterflies (based upon comparison of *Ubx* and *abd-A* RNA patterns and antibody stains) was detected with horseradish peroxidase-conjugated secondary antibodies.

FIG. 3 Repression of BX-C gene expression precedes *Dll* activation in the abdomen. *a*, Embryo at 20–25% of embryogenesis double-labelled with the rabbit antibody to *P. coenia Dll* (red) and the *Ubx/abd-A* (green) monoclonal antibody. *Dll* is expressed in the thoracic limbs but not yet activated in the abdominal limb primordia where *Ubx/abd-A* protein has disappeared. *b*, Embryo at 40% of embryogenesis double-labelled as in *b*, *Dll* (red) expression has become pronounced in the abdominal limb primordia, there may be some overlap of *Ubx/abd-A* and *Dll* expression at the expanding edges of the *Dll* domain (yellow).

METHODS. The *Ubx/abd-A* monoclonal antibody³¹ and a rabbit anti-*Dll* antibody (A. Sebring, G. Panganiban and S.C., unpublished) were detected with specific secondary antibodies and visualized by confocal microscopy.

pod or that BX-C gene expression is modulated within limb primordia as observed here in Lepidoptera. The direct analysis of myriapod homeotic genes and their expression during limb formation is the only way to resolve these fundamental issues.

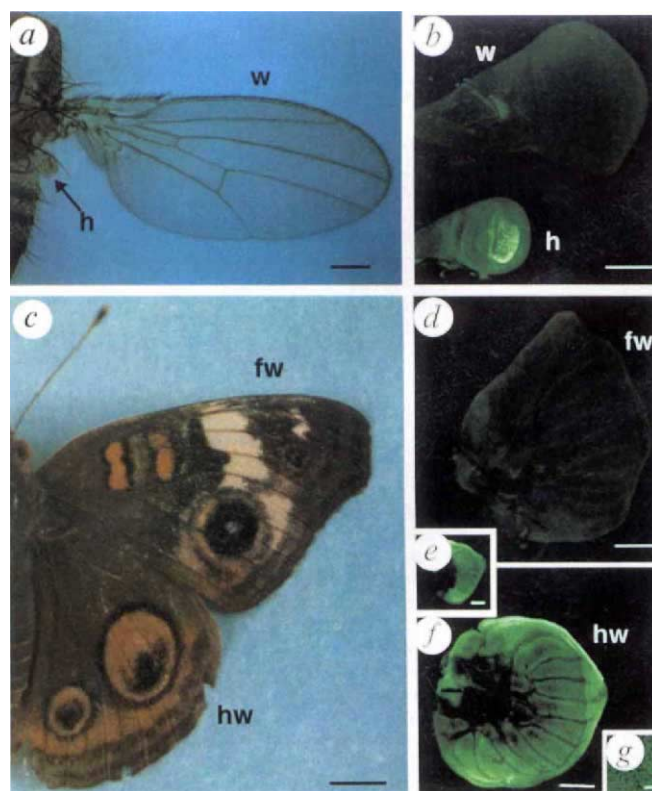
As regulatory differences in homeotic gene deployment were found to underlie differences in abdominal limb number, we also investigated whether similar regulatory differences are involved in the difference in wing number between flies and butterflies. Flies possess two pairs of flight appendages, the mesothoracic (T2) wings and the much smaller metathoracic (T3) halteres (Fig. 4*a*). *Ubx* is absent from *Drosophila* wing imaginal discs and is expressed at high levels in the portion of the dorsal metathoracic disc that will give rise to the haltere (Fig. 4*b*)^{19,20}; it is required there to suppress wing formation and to promote haltere formation^{21,22}. Butterflies possess two



large pairs of wings, forewings on T2 and hindwings on T3 (Fig. 4*c*). *Ubx* expression might be absent from developing hindwings and so permit them to grow to full-sized wings. We find, however, that *Ubx* is highly expressed in all cell nuclei in both early and late fifth instar *P. coenia* hindwing imaginal discs (Fig. 4*e-g*) but is absent from forewing imaginal discs (Fig. 4*d*), a pattern

FIG. 4 *Ubx* protein is expressed in *Drosophila* haltere and *P. coenia* hindwing imaginal discs. *a*, *Drosophila* adult hemithorax develops a large wing (w) on T2 and a small haltere (h) on T3. *b*, *Ubx* is not expressed in the wing imaginal disc (left), but is expressed at high levels on the T3 dorsal metathoracic disc (right) in cells that will form the haltere. *c*, The adult *P. coenia* develops a large forewing (fw) on T2 and a large hindwing (hw) on T3. *d*, *Ubx* is not expressed in the forewing imaginal disc. *e*, *Ubx* is expressed in the early fifth instar hindwing imaginal disc; and *f*, in the late fifth instar hindwing disc; *g*, within all cell nuclei. Scale bars: *a*, 0.25 mm; *b*, 100 μ m; *c*, 3 mm; *d*, 500 μ m; *e*, 100 μ m; *f*, 400 μ m; and *g*, 25 μ m.

METHODS. Discs were stained with the monoclonal antibody to *Ubx/abd-A* and visualized by confocal laser scanning fluorescence microscopy.



homologous to that in *Drosophila*. The presence of *Ubx* protein at high levels throughout the period during which cell number in the hindwing increases most dramatically indicates that *Ubx* is not controlling the relative size of the hindwing disc in the later larval stages, as it does in the haltere disc^{1,2,21,22}.

The regulation of *Ubx* gene expression in the posterior flight appendage primordia has been conserved between Diptera and Lepidoptera. Furthermore, the elytra (T2) and hindwings (T3) of beetles are also differentiated by the *Ubx* gene²³. So how do the posterior flight appendage primordia develop into such different structures, a large elaborately patterned hindwing or a tiny pale balancing organ, under the influence of the same homeotic gene? The most logical explanation is that the sets of downstream wing-patterning genes regulated by *Ubx* in these orders have diverged. In this view, *Ubx* operates in butterflies upon pattern-regulating genes to differentiate hindwings from forewings, and in flies upon a different set of genes to distinguish halteres from wings.

The discovery of homeotic mutants has inspired speculation about the role of homeosis²⁴ and homeotic genes^{1,2,25-27} in morphological evolution. The distinction that needs to be made, however, is whether the genesis of new homeotic genes has driven morphological change, or whether homeotic gene complexes have provided a pre-existing groundplan upon which insect segmental diversity evolved. In vertebrates new evidence has emerged that expansion of *Hox* gene clusters has occurred since the origin of primitive chordates²⁸. Evidence from phylogenetic surveys of arthropod homeotic gene complexes suggest, however, that the genetic machinery underlying segmental diversity evolved well before the morphological diversification of body plans. The results presented here demonstrate two means by which morphological diversity can be generated through changes in homeotic-regulated genetic programmes. Our study suggests that changes in the number, size and pattern of homologous structures, such as body appendages, are most likely to involve

changes in the timing and spatial regulation of existing genes, which in some cases are homeotic genes and in others, their downstream targets. □

Received 2 September; accepted 25 October 1994.

- Lewis, E. B. *Am. Zoologist* **3**, 33–56 (1963).
- Lewis, E. B. *Nature* **276**, 565–570 (1978).
- Akam, M., Dawson, I. & Tear, G. *Development* (suppl.) **104**, 123–133 (1988).
- Averof, M. & Akam, M. *Curr. Biol.* **3**, 73–78 (1993).
- Cartwright, P., Dick, M. & Buss, L. W. *Molec. Phylogenet. Evol.* **2**, 185–192 (1993).
- Wysocka-Diller, J. W., Aisemberg, G. O., Baumgarten, M., Levine, M. & Macagno, E. R. *Nature* **341**, 760–763 (1989).
- Shankland, M., Martindale, M. Q., Nardelli-Haeffliger, D., Baxter, E. & Price, D. *Development* (suppl.) **2**, 29–38 (1991).
- Dick, M. H. & Buss, L. W. *Molec. Phylogenet. Evol.* **3**, 146–158 (1994).
- Stuart, J. J., Brown, S. J., Beeman, R. W. & Denell, R. E. *Development* **117**, 233–243 (1993).
- Kelsh, R., Dawson, I. A. & Akam, M. *Development* **117**, 293–305 (1993).
- Akam, M. et al. *Development* (in the press).
- McGinnis, W. & Krumlauf, R. *Cell* **68**, 283–302 (1992).
- Karch, F. et al. *Cell* **43**, 81–96 (1985).
- Karch, F., Bender, W. & Weiffenbach, B. *Genes Dev.* **4**, 1573–1587 (1990).
- Macias, A., Casanova, J. & Morata, G. *Development* **110**, 1197–1207 (1990).
- Panganiban, G., Nagy, L. & Carroll, S. B. *Curr. Biol.* **4**, 671–675 (1994).
- Vachon, G. et al. *Cell* **71**, 437–450 (1992).
- Birket-Smith, S. J. R. *Prolegs. Legs and Wings of Insects* (Scandinavian Science, Copenhagen, 1984).
- Beachy, P. A., Helfand, S. L. & Hogness, D. S. *Nature* **313**, 545–551 (1985).
- White, R. A. H. & Wilcox, M. *Cell* **39**, 163–167 (1984).
- Morata, G. & Garcia-Bellido, A. *Wilhelm Roux Arch. dev. Biol.* **179**, 125–143 (1976).
- Struhl, G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7380–7384 (1982).
- Beeman, R., Stuart, J. J., Haas, M. S. & Denell, R. E. *Dev. Biol.* **133**, 196–209 (1989).
- Bateson, W. *Materials for the Study of Variation* (Macmillan, London, 1894).
- Goldschmidt, R. *The Material Basis of Evolution* (Yale University Press, New Haven, 1940).
- Garcia-Bellido, A. *Am. Zool.* **17**, 613–629 (1977).
- Raff, R. & Kaufman, T. *Embryos, Genes and Evolution: The Development Genetic Basis of Evolutionary Change* (Macmillan, New York, 1983).
- Garcia-Fernandez, J. & Holland, P. W. H. *Nature* **370**, 563–566 (1994).
- Sanger, F., Niden, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767–4771 (1987).
- Kelsh, R., Weinzierl, R. O. J., White, R. A. H. & Akam, M. *Dev. Genet.* **15**, 19–31 (1994).

ACKNOWLEDGEMENTS. We thank A. Laughon, P. Beachy and R. Booker for gene probes; J. Williams for the cDNA library; R. White for the *Ubx/abd-A* antibody; A. Sebring and G. Panganiban for the *P. coenia Dll* antibody; S. Paddock for assistance with microscopy and figures; L. Olds for graphics; P. Carroll, J. Langeland and A. Laughon for comments on the manuscript; and J. Wilson for its preparation. This work was supported by the Shaw Scientist Program of the Milwaukee Foundation and the Howard Hughes Medical Institute.

Distinct pathways for autocrine and paracrine Wingless signalling in *Drosophila* embryos

Joan E. Hooper

Department of Cellular and Structural Biology, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

Two secreted proteins, Wingless^{1,2} and Hedgehog^{3,4}, instruct cell fates within the segmented epidermis of *Drosophila* embryos (reviewed in ref. 5). Wingless (Wg) is expressed by the most posterior cells in each parasegment; Hedgehog (Hh) is expressed in the most anterior cells of the next parasegment. Immediately after gastrulation, the two cell types are mutually dependent^{6,7}. Local Wg signalling stabilizes Hh expression^{8–10} and local Hh signalling stabilizes Wg expression^{11,12}. Direct Wg autoregulation (autocrine signalling) is masked by its paracrine role in maintaining *hh*, which in turn maintains *wg*. I have used *zeste-white3* (*zw3*)¹³ and *patched* (*ptc*)^{11,14} mutant backgrounds to uncouple genetically this positive-feedback loop and to study autocrine Wg signalling. I report here that direct Wg autoregulation differs from Wg signalling to adjacent cells in the importance of *fused* (*fu*), *smoothed* (*sno*) and *cubitus interruptus* (*ci*) relative to *zw3* and *armadillo* (*arm*). I also find that Wg autoregulation during this early *hh*-dependent phase differs from later Wg autoregulation¹⁵ by lack of *gooseberry* (*gsh*) participation.

In embryos lacking activity of the *zw3*-encoded protein kinase, expression of Hh and the coexpressed transcription factor Engrailed (*En*) is independent of the Wg signal¹⁶. *En*+Hh expression fills the anterior half of each parasegment¹³ (E. Siegfried, personal communication), rather than being limited by the range of the Wg signal to the parasegment edge. A secondary ectopic Wg stripe is then induced at the posterior edge of the expanded *En*+Hh¹³ (Fig. 1a, b). When Wg protein expression is monitored in phenotypically null *zw3* *wg* embryos using a secretion-defective allele^{14,17,18}, Wg expression is severely reduced both in its normal and ectopic domains (Fig. 1c). This promotion of Wg expression by its own activity apparently occurs in the presence of Hh signalling. First, *En*, which is thought to regulate *hh* expression^{8–10}, is broad and robust in *zw3* *wg* embryos¹³, so *hh* expression should also be retained. Second, the residual Wg expression is limited to the edges of the expanded *En* domain, suggesting that Hh is still promoting Wg expression. I conclude that Wg activity cooperates with Hh signalling to promote Wg expression in *zw3* null embryos.

In embryos lacking activity of the *ptc*-encoded transmembrane protein^{19,20}, Wg expression no longer requires the Hh signal^{11,21}. Rather than being limited by the range of the Hh signal to the parasegment edge, Wg expression fills the posterior half of each parasegment in *ptc*⁶ and *ptc hh* embryos¹⁴ (Fig. 2a–c). In contrast, Wg activity modulates Wg expression in *ptc* mutant embryos. Instead of broadening, Wg expression fades in *wg ptc* embryos¹⁴, remaining only in occasional cells²¹ (Fig. 2d, e). *wg ptc* embryos lack *en* expression⁷, and presumably also the Hh signal. But lost *hh* expression cannot account for the reduced Wg expression in *wg ptc* embryos because Wg expression does not require Hh activity in *ptc* mutant embryos. Thus in the *ptc*

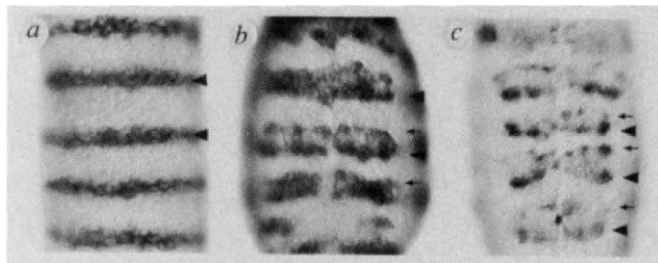


FIG. 1 Wg immunostaining in embryos lacking *zw3* activity suggests Hh-independent, Wg-dependent modulation of Wg expression. The ventral surface of a 5-h wild-type embryo (a), a 6–7-h *zw3* embryo (b) and a 6–7-h *zw3* *wg* embryo (c). The *zw3* embryos lack both maternal and zygotic *zw3* function. Anterior is up. Wg protein normally forms haloes around a narrow strip of cells, one or two cells wide (large arrowheads) marking the posterior edge of each parasegment. *zw3* embryos additionally express Wg in an ectopic stripe just posterior to the ectopic parasegment groove (small arrows). In *zw3* *wg* embryos, Wg expression is reduced and inconsistent in both its normal and ectopic domain. In some hemisegments no cells express Wg in one domain or the other. In most hemisegments occasional cells express Wg in both the normal and ectopic Wg domains.

METHODS. *zw3* null embryos came from germline clones made by γ -irradiation of *zw3*^{MA27}/*ovo*^{D1}; *wg*^{IL114}/*CyO*, *ftz-lacZ* females²⁹ and fertilized by *FM7*, *ftz-lacZ*; *wg*^{CX4}/*CyO*, *ftz-lacZ* males. The *ftz-lacZ* balancer chromosomes express β -galactosidase under control of the *fushi tarazu* promoter. Overnight embryo collections at 25 °C were processed for Wg immunohistochemistry², with or without simultaneous β -galactosidase immunohistochemical detection. Because all non-homozygous segment polarity mutant embryos express β -galactosidase, this double labelling allows unambiguous identification of the 1/8 *zw3* *wg* double null embryos within the population³⁰. The temperature-sensitive *wg*^{IL114} allele, hereafter denoted *wg*^{ts}, is phenotypically null at 25 °C (ref. 17), producing a non-secreted protein with a prolonged intracellular residence^{1,18}. It allows immunohistochemical detection of non-functional Wg. Embryo staging is based on morphology of foregut, hindgut and tracheal pits³¹. Images captured on Ektachrome 100 film were digitized using a Mira 35 Scanner and manipulated using Adobe Photoshop software.

mutant background, Hh-independent Wg signalling promotes *de novo* Wg expression away from the parasegment border and also promotes maintenance of Wg expression at the parasegment border. Although Wg function is not essential for *wg* transcription at the parasegment boundaries (for example, the residual Wg expression in *wg ptc* embryos²¹), it does upregulate its own expression by a Hh-independent mechanism, in addition to the Hh-dependent mechanism.

In *ptc* mutant embryos with low levels of Wg activity (achieved by using the temperature-sensitive *wg* allele at semi-restrictive temperature), large clusters of Wg-expressing cells replace the small clusters seen in *wg ptc* embryos (compare Fig. 2g and f). Apparently low levels of Wg activity allow the residual Wg-expressing cells in *wg ptc* mutants to nucleate larger clusters of Wg-expressing cells. The clusters usually fill the Wg-competent domain (Fig. 2h) anterior–posteriorly but not mediolaterally. Embryos simultaneously labelled for En and Wg (Fig. 2i, j) show that these clusters of Wg-expressing cells are often not adjacent to patches of En expression and so cannot be attributed to restored En expression (and Hh signal) in adjacent cells. Apparently, low Wg activity can be subthreshold for maintenance of En expression but sufficient to promote Wg expression. Mechanisms of Wg autoactivation invoking unidentified signals cannot be eliminated. The simplest explanation, however, is that secreted Wg directly activates neighbouring cells, promoting Wg expression within the Wg-competent domain.

In current models of paracrine Wg signalling, *porc* assists Wg secretion¹⁸. The Wg signal then activates Dsh, which inactivates Zw3, allowing Arm activity and En expression^{22–24}. Zw3 inactivation cannot be an intermediate step in Wg autoactivation, because Wg promotes its own expression in embryos lacking Zw3 activity (Fig. 1). *porc* and *dsh* participate in Wg autoactivation, at least in *ptc*[−] embryos because *porc ptc* or *dsh ptc* embryos lose Wg expression much as do *wg ptc* embryos (compare Fig. 3a, b to 2f). In *arm ptc* embryos, Wg is strongly expressed in about one quarter of each parasegment (Fig. 3c, d), a distinctly wider domain than wild type (Figs 1a and 2a) but distinctly narrower than *ptc* (Fig. 2c, h). Apparently, Wg expression has neither narrowed nor expanded after the blastoderm stage, suggesting that Arm is necessary for intercellular but not intra-

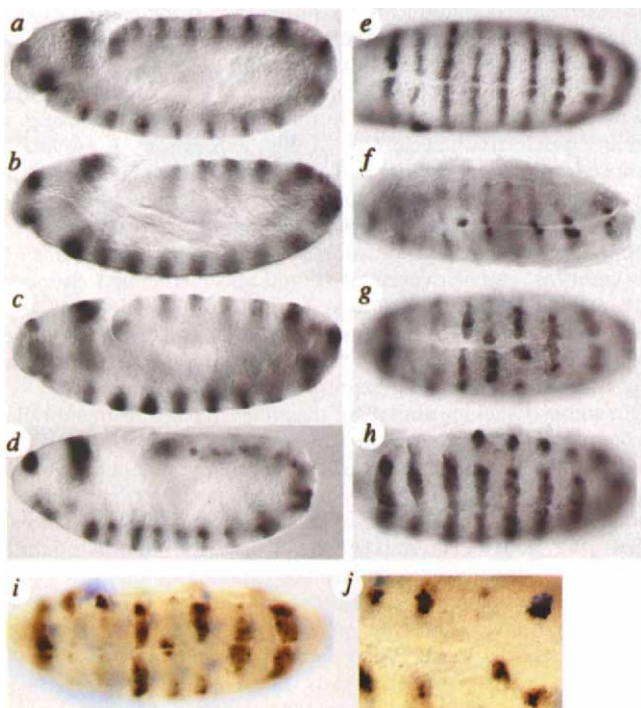


FIG. 2 Wg immunostaining in embryos lacking *ptc* activity distinguishes Hh and Wg promotion of Wg expression. Mid-sagittal optical sections (a–d), ventral (e–i), or lateral (j) surface views. Anterior is to the left (a–j) and ventral is down (a–d, j). In 5.5 h *ptc*^{6P43} embryos (b), Wg expression expands anteriorly relative to wild type (a). This broad Wg expression does not require *hh* activity, because it is unchanged in 5.5 h *ptc*^{IN108} *hh*^{13C} embryos (c). It does require *wg* activity, because Wg expression begins to fade at about 4.5 h in *wg*^{ts} *ptc*^{DF} embryos (e), is faint and irregular in all parasegments by 5 h (d), and by 5.5 to 6.5 h (f) remains only in occasional cells near the parasegment border. Stage 11 (6–7 h at 25 °C) *wg*^{ts} *ptc*^{DF} embryos raised at different temperatures titrate *wg* function (f–j). At 25 °C (f), non-functional Wg is sporadically expressed in small groups of cells, usually about 4 cells per cluster, usually near the ventral midline. At 18 °C (h), functional Wg allows Wg expression to fill the posterior half of each parasegment, marking the Wg-competent domain¹⁹. At 22 °C (g, i, j), Wg is expressed in large clusters of cells. Double labelling for En (blue) and Wg (brown) (i–j). i, Wg-independent En is strongly expressed in the nervous system (out of focus medially) and in the labial segment (out of focus anterior and lateral). The large clusters of Wg-expressing cells in right A1 and left and right T1 are not associated with ectodermal En expression, so cannot be attributed to restored En expression (and Hh signal) in adjacent cells. Dorsally (j), many Wg-expressing clusters are not associated with any En expression. **METHODS.** *hh*^{13C}, *ptc*^{6P43} and *ptc*^{IN108} are null alleles. *Df(2R)44CE* is referred to as *ptc*^{DF}. *ptc* *hh* double null embryos from *ptc*^{IN108}/*CyO*; *hh*^{13C}/*TM3*, *ftz-lacZ* parents were identified after double labelling for Wg and β -galactosidase as those lacking β -galactosidase expression while showing the *ptc* pattern of Wg expression. Wg and En double labelling was according to ref. 30 except that Wg detection used Vectastain ABC reagents.

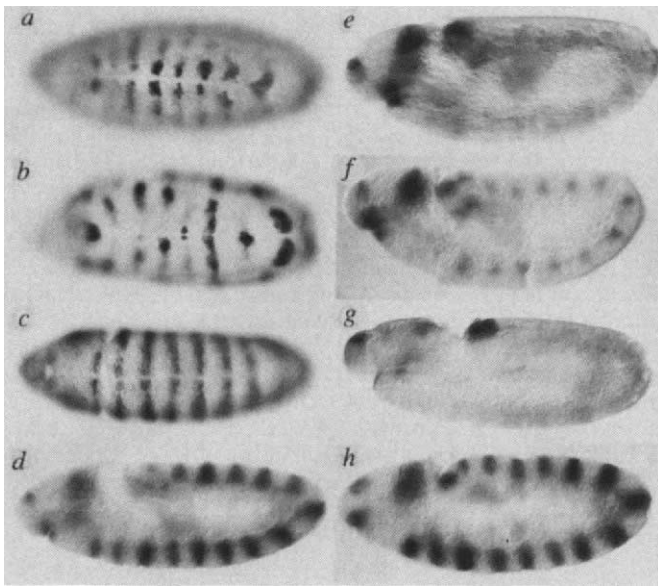


FIG. 3 Wg immunostaining in *ptc* mutant backgrounds defines requirements for *porc*, *dsh*, *arm*, *fu*, *smo*, *ci* and *gsb* in Wg autoactivation. *a*–*c*, Ventral surface or *d*–*h*, mid-sagittal optical sections of 5–8.5 h embryos. In *porc*^{P16} *ptc*^{IN108} embryos (*a*), Wg expression is limited to occasional small clusters of cells near the ventral midline (compare to Fig. 2*f*). In *dsh*^{M20} *ptc*^{IN108} embryos (*b*), Wg expression is limited to irregular discontinuous rows of cells, usually near the ventral midline. In *arm*^{YD35} *ptc*^{IN108} embryos (*c*, *d*), Wg expression neither expands anteriorly nor fades. Instead, Wg expression is consistently retained in a double row of cells in every parasegment. In *ptc*^{H84} *Ce* (*e*), *fu*⁹⁴ *ptc*^{IN108} (*f*), and *smo*^{IG26} *ptc*^{H84} (*g*) embryos, Wg expression is completely lost from the segmented ectoderm. In *ptc*^{H84} *gsb*^{DF} embryos before mid-stage 11 (*h*), Wg expression is indistinguishable from that in *ptc* or *ptc hh* embryos (Fig. 2*b*, *c*).

METHODS. *porc ptc*, *dsh ptc* and *fu ptc* embryos were derived from germ-line clones as described in Fig. 1. The *porc ptc*, *dsh ptc*, *fu ptc* and *arm ptc* crosses used balancer chromosomes marked with *ftz-lacZ* so that homozygous double mutant embryos could be distinguished within the population. *Ce*², an allele of *ci*, produces the same embryonic phenotype as a deficiency for the locus (data not shown). Cold-sensitive *smo*^{IG26} embryos were raised at 18 °C. *ptc*^{H84}, a phenotypic null, carries an enhancer trap in the *ptc* gene. *ptc* homozygotes were identified by β -galactosidase immunofluorescence. *Ce* homozygotes within the *ptc* homozygous population were distinguished by their lack of Wg immunostaining.

cellular Wg autoactivation. These zygotic null *arm ptc* embryos retain some *arm*⁺ function because germ-line *arm*⁺ is necessary for oogenesis and remains in the egg after fertilization. The unique Wg expression pattern in *arm ptc* embryos is unlikely to reflect leaky Arm function, because low levels of Wg signalling (Fig. 2*g*, *i*, *j*) generate a pattern distinct from that of *arm ptc* embryos. Arm is a component of adherens junctions²⁵ and its mammalian homologues can modulate cadherin-mediated cell adhesion^{26,27}. The *arm ptc* phenotype suggests that Arm promotes intercellular contacts that are necessary for Wg signalling to adjacent cells but not for Wg signalling within a cell.

The gene *fu*, which encodes a putative S/T kinase, acts downstream of the Hh signal to maintain Wg expression¹². *ci*[−] (ref. 28) and *smo*[−] (data not shown) embryos are very similar to *fu*[−] and *hh*[−] embryos, suggesting that they are similarly involved in Hh regulation of Wg expression. *smo*, *fu* and *ci* are more essential for Wg expression in *ptc*[−] embryos than Wg itself, because all Wg expression disappears from the segmented ectoderm in *smo ptc*, *fu ptc* and *ptc ci* mutant embryos (Fig. 3*e*–*g*). Apparently these three genes are required for Wg expression stimulated through both the Wg and Hh signals. *gsb* encodes a homeodomain-containing protein which is coexpressed with Wg and is necessary for *en* + *hh*-independent Wg autoregulation which begins

in stage 11 (ref. 15). *gsb* is not involved in Wg autoactivation in *ptc*[−] during the *hh*-dependent phase (before stage 11) discussed here (Fig. 3*h*).

Figure 4 presents a model for the autocrine and paracrine Wg signal transduction pathways. Paracrine Wg signal transduction acts through Dsh and Zw3 to activate *en* and *hh* transcription. Arm may act directly in the signal transduction pathway and/or indirectly by promoting the cell adhesion prerequisite for Wg signalling between cells. Autocrine Wg signal transduction acts through Dsh to activate Smo + Fu + Ci. In parallel, Hh from the posterior adjacent cell counteracts Ptc-mediated repression of Smo + Fu + Ci activity. The range of the Hh signal limits Wg expression to cells near the parasegment border¹². Wg autoactivation within and between Wg-expressing cells coordinates neighbouring cells within the Wg domain. The dual requirement for both the Wg and Hh signals to activate Wg expression ensures a precise and stable parasegment border despite the cell proliferation, rearrangement and variability in this developing system.

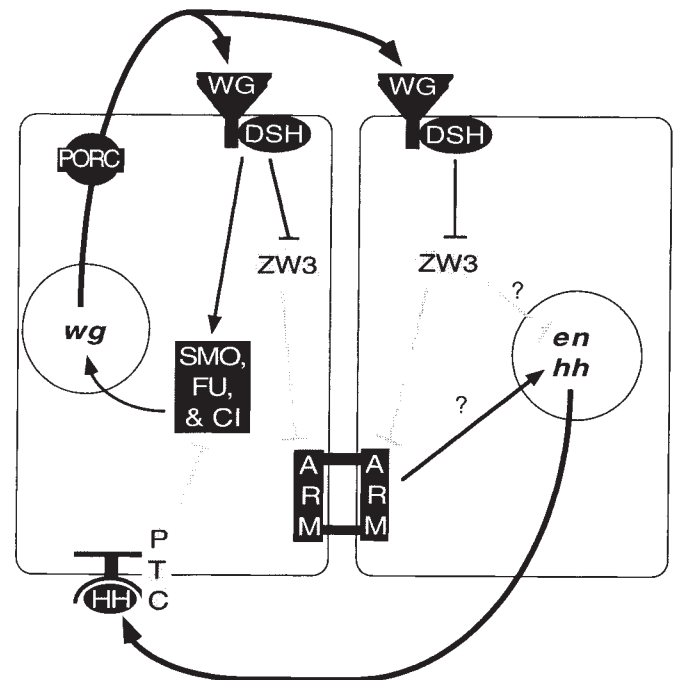


FIG. 4 Model for the autocrine and paracrine Wg signal transduction pathways. The junction between the two cells is the parasegment border. Anterior is left. Active genes, proteins and pathways are dark grey and black; inactive are light grey. Secretion of active Wg, produced in the cells at the posterior edge of each parasegment, requires *Porc*¹⁸. In both its autocrine (left cell) and paracrine (right cell) roles, Wg interacts with its (unknown) receptor to activate Dsh, which in turn inactivates Zw3 kinase (or activates a competing phosphatase)^{22,23}. This leads to Arm dephosphorylation and accumulation²⁴. Arm is dispensable for autocrine Wg signalling within a cell, but is necessary for Wg signalling between cells. I propose that, like its mammalian homologues^{26,27}, Arm at adhesive junctions²⁵ promotes intimate cell–cell contact and that this contact is prerequisite to Wg signalling between cells. Autocrine and paracrine Wg signalling pathways diverge after activation of Dsh. In autocrine signalling, Wg activates its own expression through activation of Smo + Fu + Ci. Smo + Fu + Ci activity is also regulated by Ptc-mediated repression. Transduction of the Hh signal relieves this repression^{11,12}. Reliable Wg transcription requires simultaneous activation of both pathways. Zw3 inactivation is not a direct intermediate in autocrine Wg signalling. In paracrine Wg signalling, Smo, Fu and Ci are not involved. Instead, negating Zw3 kinase activity permits Arm activation^{22,24}, transcription of *en*^{6,7,22}, and probably also *hh*. Arm may have a direct role in Wg signalling^{22–24} in addition to its role in cellular adhesion.

Note added in proof: Arm phosphorylation in *zw3* mutant embryos is reported by M. Piefer, L.-M. Pai and M. Casey in *Devl Biol.* (in the press). □

Received 23 May; accepted 11 October 1994.

- Gonzalez, F., Swales, L., Bejsovec, A., Skaer, H. & Martinez-Arias, A. *Mech. Dev.* **35**, 43–54 (1991).
- van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. *Cell* **59**, 739–749 (1989).
- Taylor, A. M., Nakano, Y., Mohler, J. & Ingham, P. W. *Mech. Dev.* **42**, 89–96 (1993).
- Tabata, T. & Kornberg, T. *Cell* **76**, 89–102 (1994).
- Perrimon, N. *Cell* **76**, 781–784 (1994).
- Martinez-Arias, A., Baker, N. E. & Ingham, P. W. *Development* **103**, 157–170 (1988).
- DiNardo, S., Sher, E., Heemskerk-Jorgens, J., Kassiss, J. & O'Farrell, P. *Nature* **332**, 604–609 (1988).
- Mohler, J. & Vani, K. *Development* **115**, 957–971 (1992).
- Tabata, T., Eaton, S. & Kornberg, T. B. *Genes Dev.* **6**, 2635–2645 (1992).
- Lee, J. J., von Kessler, D. P., Parks, S. & Beachy, P. A. *Cell* **71**, 33–50 (1992).
- Ingham, P. W., Taylor, A. M. & Nakano, Y. *Nature* **353**, 184–187 (1991).
- Ingham, P. W. *Nature* **366**, 560–562 (1993).
- Siegrfried, E., Chou, T. & Perrimon, N. *Cell* **71**, 1167–1179 (1992).
- Bejsovec, A. & Wieschaus, E. *Development* **119**, 501–517 (1993).
- Li, X., Gutjahr, T. & Noll, M. *EMBO J.* **12**, 1427–1436 (1993).
- Siegrfried, E., Perkins, L. A., Capaci, T. M. & Perrimon, N. *Nature* **345**, 825–829 (1990).
- Bejsovec, A. & Martinez-Arias, A. *Development* **113**, 471–485 (1991).
- van den Heuvel, M., Harryman-Samos, C., Klingensmith, J., Perrimon, N. & Nusse, R. *EMBO J.* **12**, 5293–5302 (1993).
- Hooper, J. E. & Scott, M. P. *Cell* **59**, 751–765 (1989).
- Nakano, Y. *et al.* *Nature* **341**, 508–513 (1989).
- Ingham, P. W. & Hidalgo, A. *Development* **117**, 283–291 (1993).
- Siegrfried, E., Wilder, E. L. & Perrimon, N. *Nature* **367**, 76–80 (1994).
- Noordermeer, J., Klingensmith, J., Perrimon, N. & Nusse, R. *Nature* **367**, 80–83 (1994).
- Piefer, M., Sweeton, D., Casey, M. & Wieschaus, E. *Development* **120**, 369–380 (1994).
- Piefer, M. *J. Cell Sci.* **105**, 993–1000 (1993).
- Hinck, L., Nelson, W. J. & Papkoff, J. *J. Cell Biol.* **124**, 729–741 (1994).
- Bradley, R. S., Cowin, P. & Brown, A. M. C. *J. Cell Biol.* **123**, 1857–1865 (1993).
- Forbes, A. J., Nakano, Y., Taylor, A. M. & Ingham, P. W. *Development* **103** (suppl.), 115–124 (1993).
- Wieschaus, E. & Noell, E. *Roux Arch. dev. Biol.* **195**, 63–73 (1986).
- Kania, M. A., Bonner, A. S., Duffy, J. B. & Gergen, J. P. *Genes Dev.* **4**, 1701–1713 (1990).
- Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* 1–227 (Springer, Berlin, Heidelberg, 1985).

ACKNOWLEDGEMENTS. I thank K. Martin, J. Ezidinma, Y. Gonzalez and M. Ayzenzon for technical assistance; R. Holmgren, J. Mohler, N. Baker, J. Nüsslein-Volhard, C. Goodman, N. Perrimon and the Bloomington Drosophila Stock Center for *Drosophila* stocks; A. Ungar, R. Holmgren, M. van den Heuvel, R. Nusse and C. Q. Doe for antibodies; T. Von Ohlen and S. Martin for comments on the manuscript. Research was supported by the NIH. (Y. Gonzalez was partially supported by a studentship from the University of Colorado Cancer Center.)

Regulation of embryonic growth and lysosomal targeting by the imprinted *Igf2/Mpr* gene

Zhao-Qi Wang, Marion R. Fung, Denise P. Barlow & Erwin F. Wagner

Research Institute for Molecular Pathology (I.M.P.), Dr Bohr-Gasse 7, A-1030 Vienna, Austria

THE receptor for insulin-like growth factor type 2, also known as the cation-independent mannose-6-phosphate receptor (*Igf2/Mpr*), is a multifunctional receptor thought to play a role in lysosomal targeting, cell growth and signal transduction^{1–8}. *Igf2/Mpr* has been mapped to the mouse *Tme*⁹ locus and shown to be an imprinted gene¹⁰, which further suggests a role in embryonic growth regulation. To define the functions of *Igf2/Mpr*, we have generated mice lacking this gene. We report here that maternal inheritance of an *Igf2/Mpr* null allele (–/+) as well as homozygosity for the inactive allele (–/–) is generally lethal at birth and mutants are about 30% larger, indicating that maternal expression of *Igf2/Mpr* is essential for late embryonic development and growth regulation. The phenotype is probably caused by an excess of Igf2 because the introduction of an *Igf2* null allele rescued the *Igf2/Mpr* mutant mice. Mutant mice also have organ and skeletal abnormalities and mis-sort mannose-6-phosphate-tagged proteins. A few (–/+) mice reactivated their paternal *Igf2/Mpr* allele in some tissues and survived to adults. But no (–/–) mice survived, indicating a role for the reactivated paternal allele in postnatal survival.

The targeting vector used to disrupt the *Igf2/Mpr* gene in embryonic stem (ES) cells is shown in Fig. 1a. Seventeen ES clones were correctly targeted at a frequency of 1/20 and one clone isolated by exposure to increased levels of G418¹¹ contained both targeted alleles (–/–, Fig. 1a). A methylation-sensitive restriction-fragment length polymorphism (RFLP) in the *Igf2/Mpr* locus¹² was used to infer the identity of the parental origin of single-targeted alleles. A paternal-specific fragment was found in 13/14 clones (data not shown) implying a preference to target the paternal chromosome. *Igf2/Mpr* messenger RNA was absent in (–/–) ES cells, indicating a complete gene inactivation, and reduced in all 14 single-targeted ES clones (Fig. 1b). If ES cells express *Igf2/Mpr* from one allele, as has been demonstrated for 13.5-day-old embryos (E13.5)¹⁰, mRNA should be abolished in maternally targeted cells and unaffected in paternally targeted clones. Because both types of targeting event caused a similar mRNA decrease (Fig. 1b, lanes –/+, +/–), this indicates that both parental alleles are expressed in ES cells in contrast to somatic cells. In this respect ES cells resemble pre-implantation embryos in which it has also been shown, by reverse-transcribed polymerase chain reaction (RT-PCR), that both parental alleles are expressed¹³.

Male chimaeric mice generated with a paternally or maternally targeted ES clone transmitted the inactivated *Igf2/Mpr* allele. Paternal transmission of the inactivated allele (pat-KO) produced viable and fertile mice heterozygous for the mutant allele (Table 1), with mRNA levels similar to wild-type littermates (Fig. 1b, lanes E16.5 +/–, +/+). In contrast, heterozygous offspring resulting from maternal transmission of the inactivated allele (mat-KO) were viable *in utero* but mostly moribund at birth, or dead within 2 weeks (Table 1), and expressed very low levels of mRNA from the paternal chromosome (Fig. 1b, lane E16.5 –/+). A small number of long-term surviving, but so far infertile, mat-KO heterozygotes were obtained on a C57BL6/129Sv background. However, none survived on an inbred 129/Sv background (Table 1). The phenotype of offspring homozygous for the mutated allele (–/–) obtained from heterozygous matings, was similar to that of mat-KO heterozygotes. However, no long-term (–/–) survivors have yet been obtained in 304 offspring analysed (Table 1). The survival of some mat-KO (–/+) mice, together with the absence of any (–/–) long-term survivors, suggests that the paternal chromosome known to be repressed at E13.5, may be reactivated after birth. Reactivation was examined in one mat-KO long-term survivor which showed expression of paternal *Igf2/Mpr* in lung, kidney and brain (Fig. 1b, lanes Lu, Ki, Br). These data identify a critical role for *Igf2/Mpr* in late embryonic and neonatal survival.

Mat-KO (–/+) and homozygous mutant (–/–) embryos develop to term but the majority cannot initiate or sustain respiration and die at birth. Additionally, all moribund mat-KO and (–/–) E18.5 fetuses were ~30% heavier in dry weight than wild-type littermates (Fig. 2a, b), have a tail bend (Fig. 2a) and a single extra toe on the post-axial side of fore and hindfeet (Fig. 2c, d). In moribund E18.5 mutants, the limb phalanges and tail vertebrae had ossified to a stage characteristic of later post-natal stages¹⁴ (Fig. 2d–f). Moribund mutants also showed skeletal anomalies, including ossified extra toes, fused third and fourth sternum ossification centres (Fig. 2d–h), and a split sternum similar to, but less severe than that described for the *Hoxb-4* null mutation¹⁵ (Fig. 2h). Histological analysis of several E18.5 moribund mat-KO and (–/–) mice revealed abnormalities in the lungs and heart. The lung appeared retarded and alveoli poorly formed (not shown). The heart ventricular myocardium was disorganized and the heart enlarged to between two and five times that of controls (not shown). The few surviving mat-KO mice show a smaller increase in birthweight than moribund littermates and have tail bends that are seen at birth but straighten with age. Two surviving (–/+) adult females had abnormally large uteri connecting to a fused vagina rendering them infertile.

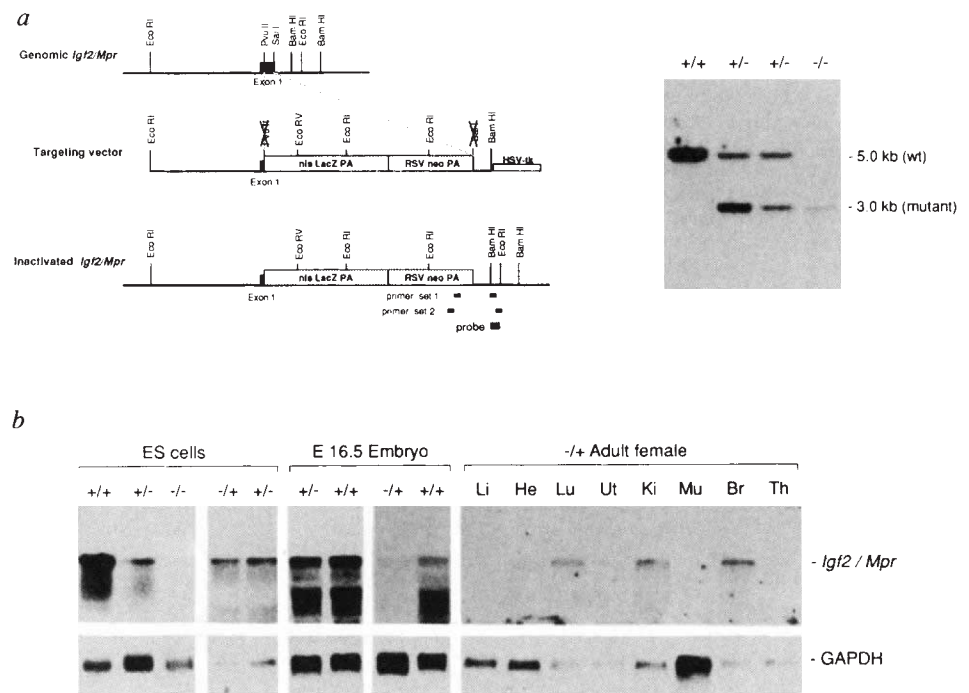
TABLE 1 Parental transmission of the *Igf2/Mpr* null allele

Mating	Age	(+/+)	(+/-) Paternal inheritance	(-/-) Maternal inheritance	(-/-)	Total
Paternal transmission (+/+) × (+/-)	Adult	119	103	—	—	222
Maternal transmission (+/-) × (+/+)	E14.5–18.5	67	—	67	—	134
	Adult (BL6/129)	168	—	4	—	172
	Adult (129/129)	109	—	0	—	109
Heterozygous cross (+/-) × (+/-)	E14.5–18.5	36	— 73 —	—	25	134
	Adult (BL6/129)	163	140	1	0	304

Chimaeras that transmitted the inactive allele were obtained from the maternally targeted D3-a3, and paternally targeted D3-a13 ES clones. *Igf2/Mpr* genotype was analysed by Southern blots (Fig. 1). Offspring were generated from heterozygote crosses (+/+ × +/-) or (+/- × +/+) and heterozygous intercrosses (+/- × +/-). Because the *Igf2/Mpr* gene is maternally expressed in the late embryo¹⁰, maternal inheritance of a targeted allele (-/+) exhibits a mutant phenotype, whereas paternal inheritance generates phenotypically normal (+/-) heterozygous offspring. In the intercross, mutant (-/+) offspring were distinguished in adult stages by the external phenotype shown in Fig. 2a, c. BL6, C57BL/6; 129, 129/Sv.

FIG. 1 a, The targeting vector contains 3.9 kilobases (kb) of 5' genomic DNA and 0.6 kb of 3' sequence. The neomycin expression cassette is driven by an RSV promoter and contained a polyadenylation site. The pGNA vector²⁶ was modified to include a nuclear localization signal (nls) fused to the LacZ reporter gene (nlsGNA) that was inserted into the *Igf2/Mpr* gene replacing a 96 base pair (bp) *PvuII*–*Sall* fragment from the first exon. Southern blot analysis using *EcoRI* digestion and an external probe detects a 5.0 kb wild-type allele and a 3.0 kb fragment specific for the disrupted allele. Wild-type (+/+), two single-targeted (+/-) and one double-targeted (-/-) ES cell clone are shown. b, Expression of *Igf2/Mpr* mRNA and a GAPDH loading control in ES cells, embryonic and adult tissues. The minimal signal present in (-/+) embryonic tissue demonstrates that the paternal chromosome is repressed but not completely silent. Li, liver; He, heart; Lu, lung; Ut, uterus; Ki, kidney; Mu, muscle; Br, brain; Th, thymus.

METHODS. A 5.0 kb *EcoRI* fragment of 5' sequences of the *Igf2/Mpr* gene derived from a C57BL/6 genomic library was isolated. A 3.9 kb *EcoRI*–*PvuII* region that included the *Igf2/Mpr* promoter and the first 4 codons was inserted at the upstream *XmnI* site of pnsGNA and a 1.8 kb *HSV-tk* gene with its own promoter and a 0.6 kb *Sall*–*BamHI* 3' region of *Igf2/Mpr* were inserted into the downstream *SmaI* site of pnsGNA. In this construct the LacZ gene is fused in-frame with the first exon and LacZ expression should be regulated by the *Igf2/Mpr* gene promoter. The targeting vector was linearized with *XmnI* and electroporated



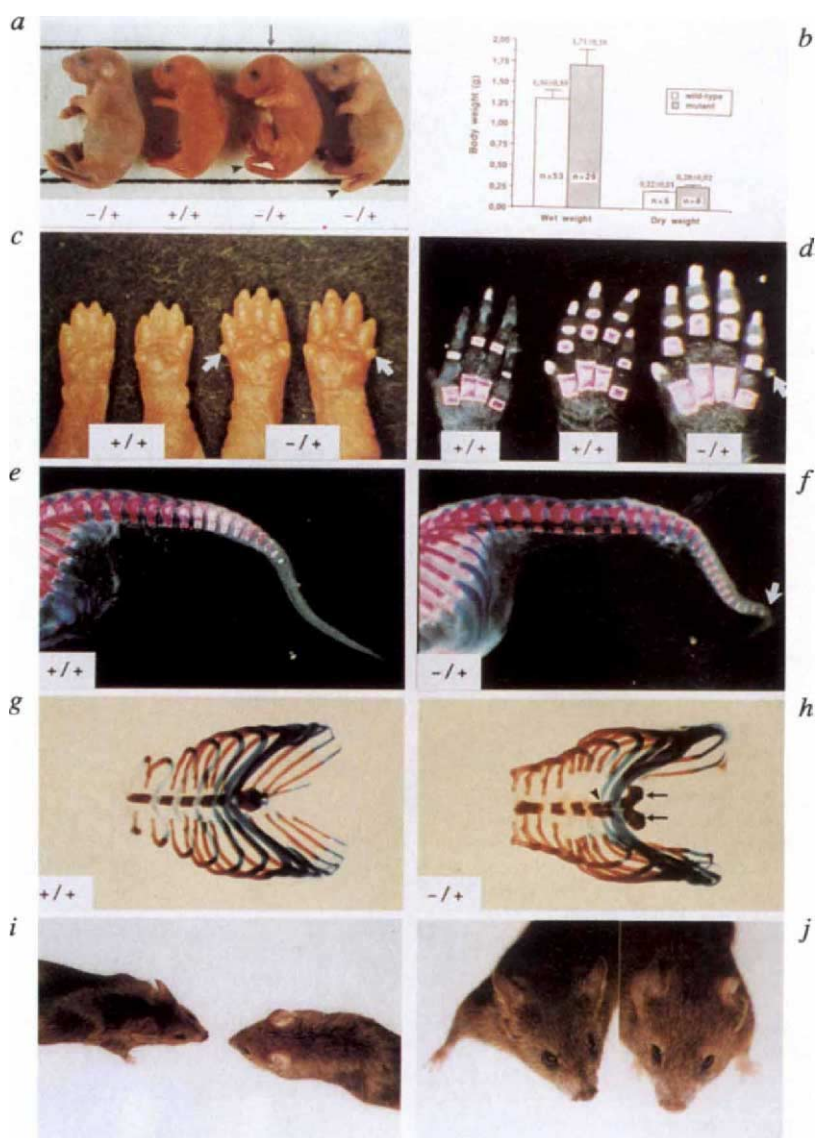
(not shown). Surviving mice also displayed a facial dysmorphism (Fig. 2i, j) that has similarities to that found in β -glucuronidase-deficient mice¹⁶, and may be analogous to those reported in humans with mucopolysaccharidoses¹⁷.

Igf2/Mpr targets the majority of its mannose-6-phosphate-tagged ligands to lysosomes^{1,2} and is expected to play a major role in intracellular sorting of newly synthesized enzymes and in endocytosis of extracellular enzymes. Intracellular sorting was examined in primary embryo fibroblasts isolated from E14.5 embryos. Steady-state levels of four out of five lysosomal enzymes targeted through a mannose-6-phosphate tag were

reduced by 30–80% in (-/+) and (-/-) embryo fibroblasts lacking *Igf2/Mpr* (Fig. 3a). These reductions were not seen in cell extracts from mutant E14.5 embryos or in hearts and livers of E18.5 embryos, probably because these extracts contain intracellular and extracellular enzyme pools. However, enzyme levels in amniotic fluid showed a three to eight times increase in E14.5 and E16.5 mutants. Endocytosis of extracellular ligands was examined in two ways. First, by internalization of *Igf2/Mpr* antiserum, a process that did not occur in (-/-) E14.5 embryonic fibroblasts (Fig. 3b). However, about 1/500 (-/+) embryonic fibroblasts internalized antiserum, suggesting that

FIG. 2 Phenotype of E18.5 *Igf2/Mpr* mutant mice. **a**, The external appearance of moribund ($-/+$), wild-type littermate ($+/+$) and a surviving ($-/+$) fetus is shown. Moribund *Igf2/Mpr* null mice are larger in size and have a tail bend (arrowhead), the surviving ($-/+$) mouse (arrow) has a slight increase in size. A tail bend has also been documented for the *Tme* mutant¹⁰. **b**, The body weight of mutant newborns is clearly different from that of wild-type littermates; however, the water content of mutants is similar to wild-type littermates (83.1% versus 83.6%), indicating the absence of oedema. Increased body weight has also been described for the *Tme* mutant which has been ascribed to oedema¹⁰; we therefore measured the dry weight of three E15.5–E17.5 litters derived from ($T^{Hb}/+ \times +/+$) crosses. The weight of the *Tme* mutant embryos was about 24% greater than wild-type littermates; however, the standard deviation overlapped between the two groups. The water content of *Tme* mutants was increased by 3–5% compared to wild-type littermates. *Igf2/Mpr* mutant mice have an extra toe on the post-axial side of the forefeet (arrow in **c**) or hindfeet (not shown), as has also been documented for the *Tme* mutant¹⁰. Skeletal analysis showed the extra toe is ossified (arrow in **d**). The ossification of digits was more advanced in mutants, despite the variation that can exist in wild-type littermates (**d**). Ossification of the vertebral column (**e**, **f**) was also more advanced in mutants than in wild-type littermates, but no vertebral malformations were associated with the tail bend which occurred distal to the onset of ossification (arrow). Mutant mice exhibit a split sternum (arrow in **h**) and fused 3rd and 4th sternum ossification centres (arrowhead in **h**) compared to a wild-type littermate (**g**). **i**, **j**, Show the facial appearance of an adult littermate (left in both photographs) and a rare surviving *Igf2/Mpr* ($-/+$) adult (right in both photographs). The mutants have a shortened facial long axis.

METHODS. E18.5 fetuses were prepared for skeletal analysis as described²⁸ and for histological examination as described²⁷ (six mutant and wild-type embryos were examined at E18.5 for skeletal changes and for histology). Dry weight was measured after 48–96 h at 50 °C.



sporadic expression of the paternal allele continued beyond E14.5. Second, by internalization of human β -glucuronidase, a process that also did not occur in ($-/-$) cells but was detectable in ($+/+$) cells (mean enzyme activity of two wild-type cell lines was 719 ± 69 units, see Fig. 3 legend). Taken together, these data indicate that although *Igf2/Mpr* is essential for endocytosis of extracellular ligands, its role in intracellular sorting is partially compensated by alternative pathways such as the *MPR46* receptor^{1,2}. However, the exact contribution of the enzyme sorting defect to the phenotypes depicted in Fig. 2a is not yet clear.

The phenotype of mice lacking *Igf2/Mpr* greatly overlaps that described for the *Tme* mutation⁸. Both mutations produce a complex phenotype on maternal transmittance, featuring increased body size and dry weight, polydactyly, bent tails and a late embryonic/neonatal lethality dependent on the genetic background. This similarity argues in favour of *Igf2/Mpr* being the *Tme* gene, but contrasts with the claim, based on the rescue of *Tme* in mice with a mixed genetic background, that *Tme* and *Igf2/Mpr* are distinct loci¹⁸. Our hypothesis is supported by rescue experiments using *Igf2* null mice which suggest that an excess of the ligand *Igf2* is the primary cause of death in the *Tme* mutant¹⁹. We have directly tested this by breeding mice lacking both *Igf2/Mpr* and *Igf2*²⁰ and preliminary experiments

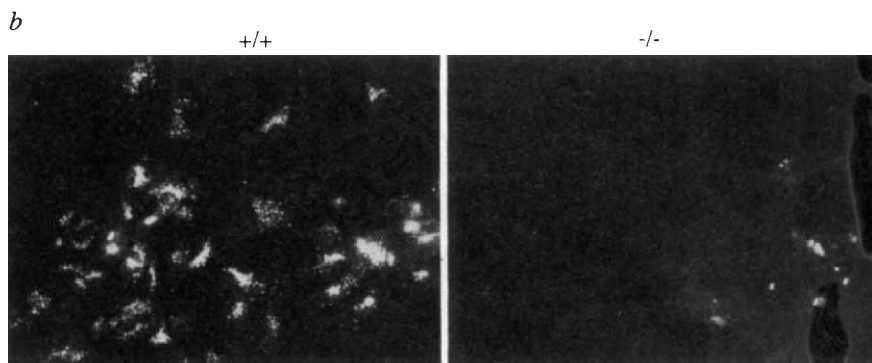
show that the absence of *Igf2* can completely rescue the phenotypes depicted in Fig. 2a–f in six live born offspring from three litters (data not shown). The complex embryonic phenotype displayed by the *Igf2/Mpr* mutant could result from high serum concentrations of active ligands such as *Igf2*³ and/or accumulation of inactive ligands such as latent-TGF- β ⁵ and proliferin⁶. It is possible, but unlikely, that reduced steady-state levels of lysosomal enzymes contribute to the embryonic phenotype, because similar defects are present in *Mpr46* null mice with no obvious effect on viability^{21,22}. Biallelic expression of *Igf2/Mpr* in ES cells, together with reactivation of the paternal allele observed in mat-KO surviving mice, suggests that monoallelic expression of imprinted genes is regulated in a stage- and cell-type-specific manner. The *Igf2/Mpr* mutant mouse should allow a dissection of the mechanisms of allele-specific repression and reactivation.

These results indicate that the primary role of the imprinted *Igf2/Mpr* gene may vary between the adult and embryo. In the adult the receptor may act primarily to maintain enzyme levels, and thus the activity of lysosomes. In the embryo, however, *Igf2/Mpr* primarily inhibits growth. This, together with the demonstration that the imprinted *Igf2* gene promotes embryonic growth²³ and the identification of anomalies of imprinted gene

FIG. 3 a, Lysosomal enzyme levels in four wild-type (+/+ and +/-) and two *Igf2/Mpr* null (-/+ and -/-) primary embryo fibroblasts (MEF) and in amniotic fluid. The levels of six lysosomal enzymes (acid phosphatase is targeted independently of M6P and serves as a control) were determined. Steady-state levels in MEFs were reduced in mutant cell lines by two- to fourfold, except for β -galactosidase. Levels in amniotic fluid were elevated in E14.5 and E16.5 embryos devoid of *Igf2/Mpr*. b, Uptake of *Igf2/Mpr* antiserum in wild-type (clone 16 +/+) and mutant (clone 17 -/-) MEF lines. Cell lines expressing *Igf2/Mpr* (+/+) internalized the antiserum but negative cells (-/-) did not (the spots in the bottom left of the field of view are nonspecific). nd, Not done.

METHODS. a, MEF lines derived from E14.5 wild-type embryos were grown in heat-inactivated medium. Cells were scraped into 0.9% NaCl, frozen and thawed once and the supernatant assayed. Embryonic tissue was homogenized to a single cell suspension in 0.9% NaCl and similarly processed. Amniotic fluid was assayed directly. Numbers are enzyme units (one unit equals 1.0 nM of 4-methylumbelliferone per h per mg protein at 37 °C), mean $\pm$ s.e.m. Assays measure, by fluorimetry, the release of 4-methylumbelliferone from specific substrates²⁹. b, Cells were incubated with rat *Igf2/Mpr* antiserum, fixed, permeabilized and stained with FITC-labelled goat anti-rabbit IgG. Uptake of human β -glucuronidase was assayed on the same clones (see text). Human β -glucuronidase was produced by infecting mouse MPSVII cells with supernatant from the ψ CRIP-GLU-3 cell line which produced high titres of recombinant retrovirus carrying

a	α -Mannosidase	β -Glucuronidase	α -Galactosidase	β -Galactosidase	β -Hexosaminidase	Acid phosphatase
MEF (E14.5)						
mutants (n=2)	8.9 $\pm$ 0.5	46.5 $\pm$ 10	336.2 $\pm$ 41	497 $\pm$ 93	1,894 $\pm$ 177	2,226 $\pm$ 230
wild-type (n=4)	42.8 $\pm$ 13	292.5 $\pm$ 66	496.9 $\pm$ 117	600 $\pm$ 93	5,138 $\pm$ 152	2,124 $\pm$ 152
Amniotic fluid (E14.5)						
mutants (n=3)	92 $\pm$ 27	293 $\pm$ 56	0	297 $\pm$ 76	15,513 $\pm$ 3,542	0
wild-type (n=3)	11 $\pm$ 4	45 $\pm$ 28		95 $\pm$ 84	5,579 $\pm$ 2,598	
Amniotic fluid (E16.5)						
mutants (n=4)	1,090 $\pm$ 734	20,769 $\pm$ 10,212	nd	12,803 $\pm$ 7,059	530,028 $\pm$ 290,734	nd
wild-type (n=4)	0	5,001 $\pm$ 2,119		2,131 $\pm$ 880	133,266 $\pm$ 103,065	



expression during tumour development^{24,25}, argues that imprinted genes have fundamental roles in the control of cell growth in development and oncogenesis. □

Received 31 August; accepted 2 November 1994.

1. von Figura, K. *Curr. Opin. Cell Biol.* **3**, 642–646 (1992).
2. Kornfeld, S. A. *Rev. Biochem.* **61**, 307–330 (1992).
3. Kyle, J. W., Nolan, C. M., Oshima, A. & Sly, W. S. *J. Biol. Chem.* **263**, 16230–16235 (1988).
4. Oka, Y., Rozek, L. M. & Czech, M. P. *J. Biol. Chem.* **260**, 9435–9442 (1985).
5. Dennis, P. A. & Rifkin, D. B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 580–584 (1991).
6. Lee, S.-J. & Nathans, D. *J. Biol. Chem.* **263**, 3521–3527 (1988).
7. Herzog, V., Neumüller, W. & Holzman, B. *EMBO J.* **6**, 555–560 (1987).
8. Murayama, Y. et al. *J. Biol. Chem.* **265**, 17456–17462 (1990).
9. Johnson, D. R. *Genetics* **76**, 795–805 (1974).
10. Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K. & Schweifer, N. *Nature* **349**, 84–87 (1991).
11. Mortensen, R. M., Conner, D. A., Chao, S., Geisterfer-Lawrence, A. A. T. & Seidman, J. G. *Molec. cell. Biol.* **12**, 2391–2395 (1992).
12. Stöger, R. et al. *Cell* **73**, 61–71 (1993).
13. Latham, K. E., Doherty, A. S., Scott, C. D. & Schultz, R. M. *Genes Dev.* **8**, 290–299 (1994).
14. Kaufman, M. H. in *The Atlas of Mouse Development* (ed. Kaufman, M. H.) (Academic, London, 1992).
15. Ramirez-Solis, R., Zheng, H., Whiting, J., Krumlauf, R. & Bradley, A. *Cell* **73**, 279–294 (1993).
16. Vogler, C. et al. *Am. J. Path.* **136**, 2207–2217 (1990).
17. Nolan, C. M. & Sly, W. S. in *Metabolic Basis of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 1589–1601 (McGraw-Hill, New York, 1989).
18. Forejt, J. & Gregorova, S. *Cell* **70**, 443–450 (1992).
19. Filson, A., Louvi, A., Efstradiadis, A. & Robertson, E. J. *Development* **118**, 731–736 (1993).
20. DeChiara, T. M., Efstradiadis, A. & Robertson, E. J. *Nature* **345**, 78–80 (1990).
21. Köster, A. et al. *EMBO J.* **12**, 5219–5223 (1993).
22. Ludwig, T. et al. *EMBO J.* **12**, 5225–5235 (1993).
23. DeChiara, T. M., Robertson, E. J. & Efstradiadis, A. *Cell* **64**, 849–859 (1990).
24. Feinberg, A. P. *Nature Genet.* **4**, 110–113 (1993).
25. Christofori, G., Naik, P. & Hanahan, D. *Nature* **369**, 414–418 (1994).
26. Le Mouellie, H., Lallemand, Y. & Brulet, P. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4712–4716 (1990).
27. Wang, Z.-Q. et al. *Nature* **360**, 741–745 (1992).
28. McLeod, M. J. S. *Teratology* **22**, 299–301 (1980).
29. Moullier, P., Marechal, V., Danos, O. & Heard, J. M. *Transplantation* **56**, 427–432 (1993).
30. Frankel, H. A., Glaser, J. H. & Sly, W. S. *Pediat. Res.* **11**, 811–816 (1977).

ACKNOWLEDGEMENTS. We thank L. Stingl for technical assistance, N. Howells for animal husbandry, A. Aguzzi for advice and help with histology, K. Nolan for advice and discussions on lysosome assays, K. von Figura for the rat *Igf2/Mpr* antiserum, O. Danos and N. Naffakh for providing the material and advice on the human β -glucuronidase uptake experiment, A. Efstradiadis for supplying the *Igf2* knockout mice, and A. Grigoriadis, A. Aguzzi, M. L. Birmstiel, K. von Figura and B. Hogan for reading the manuscript. This research was supported in part by the Austrian Industrial Research Promotion Fund.

the human β -glucuronidase cDNA regulated by a mouse PGK promoter (cells were provided by O. Danos and N. Naffakh). MEF cells were incubated for 9 h with the heat-stable human β -glucuronidase enzyme and extracts heat-treated to inactivate the mouse enzyme before assay³⁰.

A topoisomerase II-dependent G2 cycle checkpoint in mammalian cells

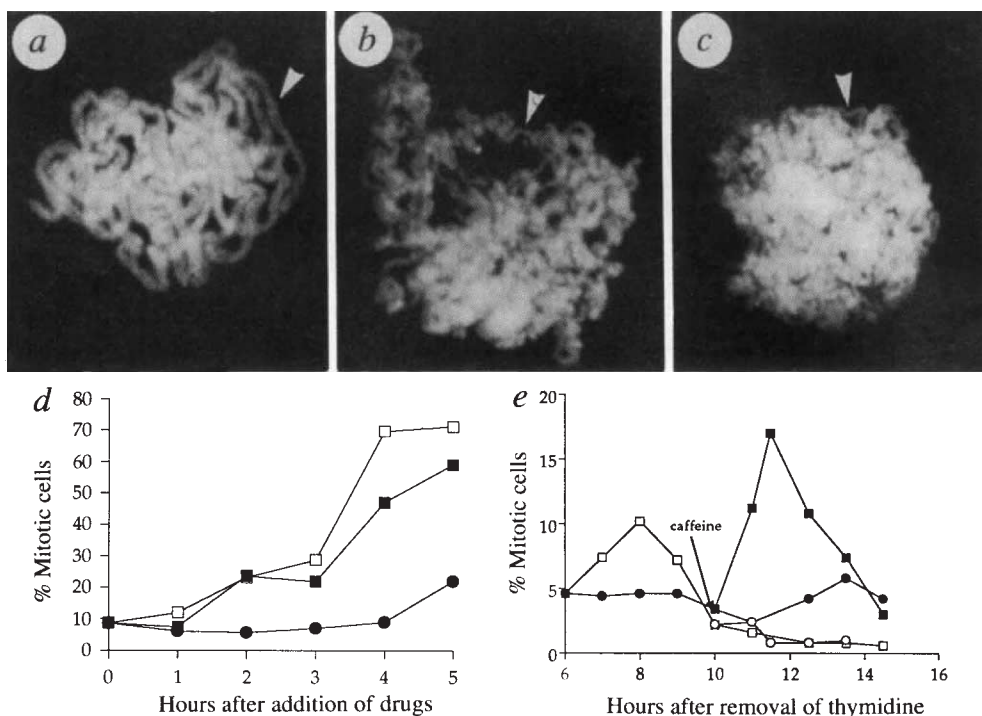
C. Stephen Downes, Duncan J. Clarke, Ann M. Mullinger, Juan F. Giménez-Abián, Andrew M. Creighton* & Robert T. Johnson

CRC Mammalian Cell DNA Repair Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

* Medicinal Chemistry Laboratory, Department of Reproductive Physiology, St Bartholomew's Hospital, London EC1A 7BE, UK

THE enzyme DNA topoisomerase II, which removes the catenations formed between the DNA molecules of sister chromatids during replication¹ and is a structural component of chromosome cores², is needed for chromosome condensation in yeast³ and in *Xenopus* extracts^{4–6}. Inhibitors of topoisomerase II arrest mammalian cells before mitosis in the G2 phase of the cell cycle⁷, but also produce DNA damage, which causes arrest through established checkpoint controls⁸. It is open to question whether cells need topoisomerase II to leave G2, or control late-cycle progression in response to its activity. Bisdioxopiperazines are topoisomerase II inhibitors that act without producing direct DNA damage⁹; the most potent, ICRF-193, blocks mammalian entry into but not exit from mitosis. Here we show that checkpoint-evading agents such as caffeine override this block to produce abortively condensed chromosomes, indicating that topoisomerase II is needed for complete condensation. We find that exit from G2 is regulated by a catenation-sensitive checkpoint mechanism which is distinct from the G2-damage checkpoint.

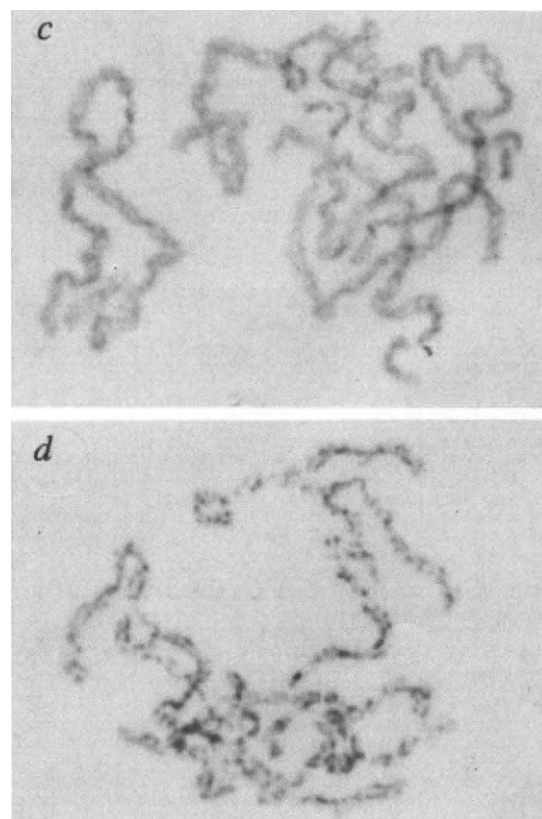
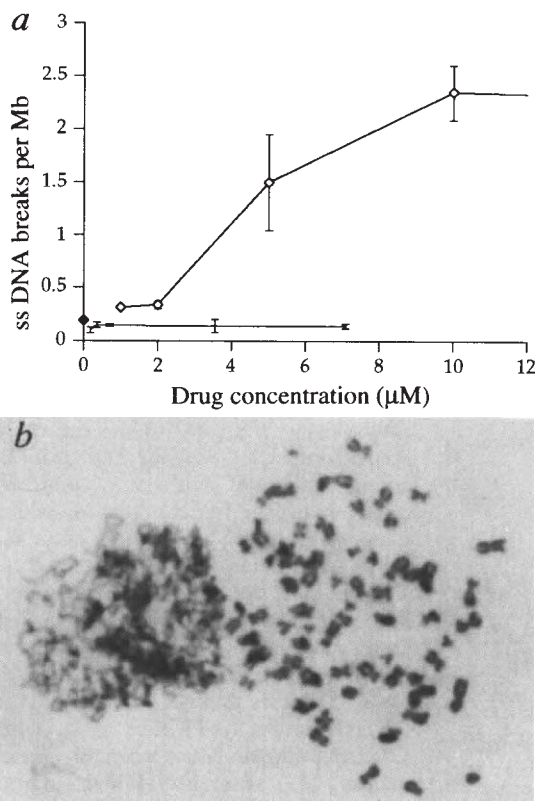
FIG. 1 Pseudo-mitotic chromosome condensation in the absence of topoisomerase II activity. **a-c**, Indian muntjac chromosomes (convenient for their large size) from DM87 cells given 2 mM caffeine and 2 $\mu\text{g ml}^{-1}$ ICRF-193 for 1, 4 and 6 h; cells were in late G2, G2/S and S phase at time of addition of drugs, giving rise to progressively less condensed chromosomes. Magnification, $\times 986$; arrows indicate well defined regions of chromatids. Similar structures are seen on treatment of cells with ICRF-193 and 1 mM 6-dimethylaminopurine or 100 nM okadaic acid (not shown). **d**, Accumulation of HeLa cells in mitosis or pseudo-mitosis when given 0.04 $\mu\text{g ml}^{-1}$ nocodazole to arrest in metaphase ($\square$), nocodazole and ICRF-193 ($\bullet$), or nocodazole, ICRF-193 and caffeine ($\blacksquare$). Cells were presynchronized in S phase with a thymidine block and drugs added 5 h after release. Other mammalian cell lines (CHO, PtK2 rat kangaroo, SVM Indian muntjac, SV-MRC5 transformed human and MSU1.1 immortalized human²⁵) respond similarly (not shown). Caffeine alone fails to affect entry into mitosis in any line tested. **e**, Release of MSU1.1 cells from ICRF-193-imposed G2 arrest by caffeine, in the absence of microtubule antagonists; addition of caffeine to cells delayed in G2 produces a wave of pseudomitotic cells with imperfectly condensed chromosomes. Accumulation of cells in mitosis or pseudo-mitosis following release from a thymidine block and given: ICRF-193 after 6 h and caffeine added after a further 4 h ($\blacksquare$), or ICRF-193 after 6 h ($\bullet$), or caffeine after 10 h ($\circ$), or no further treatment ($\square$).



METHODS. Cells were grown in monolayer, and chromosome preparations,

made at intervals from hypotonically swollen cells¹⁹, were stained with propidium iodide for confocal microscopy (**a-c**) or crystal violet for scoring by bright-field microscopy (**d, e**). At least 500 cells were scored for each data point. 2 $\mu\text{g ml}^{-1}$ ICRF-193 ($\sim 7 \mu\text{M}$) is a concentration high enough to inhibit mammalian mitotic topoisomerase II completely, as is 60 μM etoposide¹⁶. Proliferating unsynchronized cultures were used for **a-c**.

FIG. 2 Differences between ICRF-193-induced and etoposide-induced effects. **a**, DNA strand breaks induced in G2 HeLa cells treated with various concentrations of etoposide ($\diamond$) or ICRF-193 ($\circ$); background ($\blacklozenge$). **b**, Premature chromosome condensation in G2 HeLa cells fused with mitotic HeLa in the presence of 2 $\mu\text{g ml}^{-1}$ ICRF-193. Such premature chromosome condensation cannot be induced in G2 HeLa fused with mitotic HeLa in the presence of 60 μM etoposide, unless caffeine is also present (not shown). **c**, Indian muntjac chromosomes from SVM cells condensed by caffeine in the presence of 2 $\mu\text{g ml}^{-1}$ ICRF-193. **d**, Breaks and translocations in SVM chromosomes condensed by caffeine in the presence of 60 μM etoposide. METHODS. Strand breaks were assayed in HeLa cells prelabelled with ^3H -thymidine, synchronized by a double thymidine block followed by nocodazole depletion of mitotic cells²⁶, using a version of the alkaline unwinding/hydroxylapatite chromatography technique²⁷, calibrated against breaks induced by ^{137}Cs γ -radiation as described²⁸. HeLa cells for fusion were



synchronized in G2 by a double-thymidine block, and in mitosis by a thymidine/nitrous oxide block, as described²⁶. Premature chromosome condensation was induced by fusion of G2 and mitotic cells with Sendai virus²⁹.

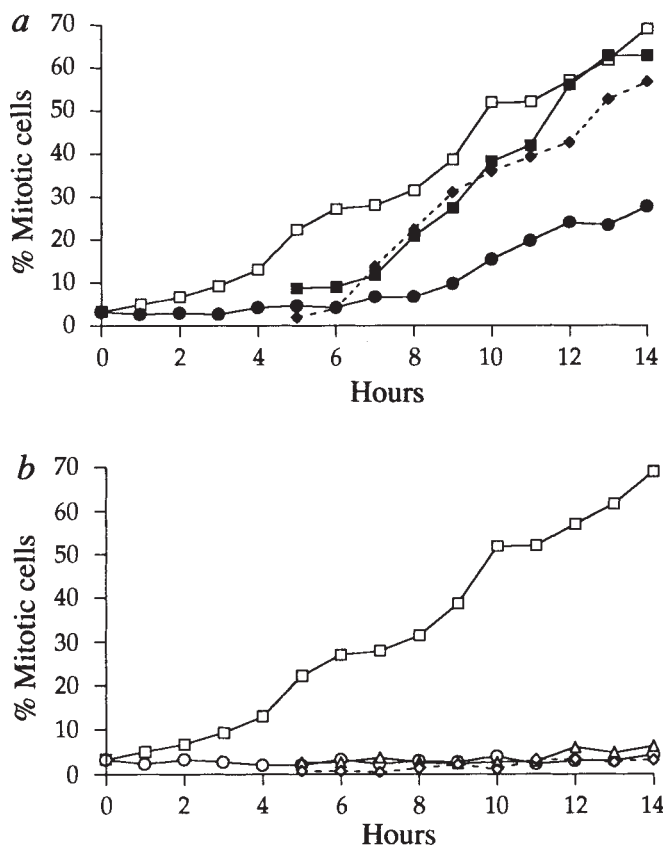
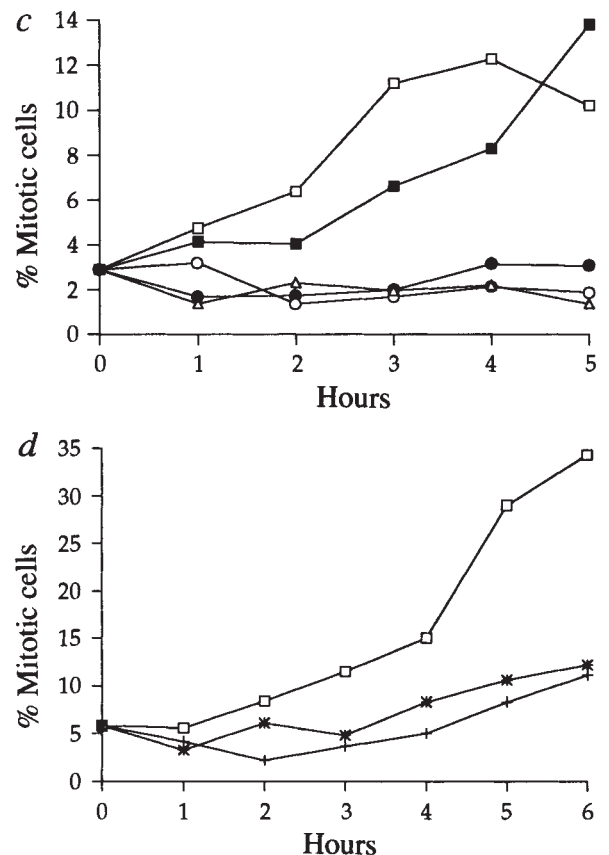


FIG. 3 Differences in cycle progression after treatment with ICRF-193 and etoposide. *a*, Recovery of progression into mitosis in HeLa after prolonged ICRF-193 arrest; cells accumulated with nocodazole (□) or 2 $\mu\text{g ml}^{-1}$ ICRF-193 plus nocodazole (●), or given 4 h with ICRF-193 and nocodazole, and then either given 2 mM caffeine (■), or washed and returned to normal medium with nocodazole (◆). Washing allows progress into normal mitosis; caffeine induces imperfectly condensed pseudomitosis. *b*, Failure of HeLa to recover after prolonged etoposide arrest; cells accumulated with nocodazole (□) or 60 μM etoposide plus nocodazole (○), or given 4 h with etoposide and nocodazole and then either given 2 mM caffeine (△) or washed and given nocodazole (◇).

To determine whether the G2 delay induced by topoisomerase II inhibitors is due to a physical inability to perform pre-mitotic condensation in the presence of DNA catenation, or to a checkpoint control system sensitive to catenation, we have combined two approaches. First, we inhibit topoisomerase II and then give agents that allow bypass of replicative and post-replicative checkpoints: caffeine¹⁰, and also kinase and phosphatase inhibitors^{11,12}. Second, we use a new topoisomerase II inhibitor: one of the ICRF series of bisdioxopiperazines, antitumour agents that delay passage of mammalian cells into mitosis¹³⁻¹⁵. These are powerful inhibitors of topoisomerase II; and unlike other inhibitors, which arrest the enzyme at its transition state after DNA-strand breakage, bisdioxopiperazines trap the enzyme in the form of a closed protein clamp⁹, with no associated DNA damage¹⁶⁻¹⁸.

In mammalian cells ICRF-193 prevents progression from G2 to mitosis for several hours, but agents that promote checkpoint evasion allow cells to attempt chromosome condensation in the presence of ICRF-193 without delay (Fig. 1). We have monitored progression into mitosis by accumulation with the microtubule antagonist nocodazole; however, an ICRF-193-induced G2 delay that can be overridden by caffeine, producing partly condensed chromosomes, can also be demonstrated without nocodazole (Fig. 1e).

Such partial condensation indicates that topoisomerase II function is not necessary for exit from G2, and that therefore



c, Caffeine overrides G2 delay imposed by ICRF-193, but not by etoposide, in DM87 cells; mitotic or pseudo-mitotic accumulation of cells given 0.04 $\mu\text{g ml}^{-1}$ nocodazole (□), nocodazole and 2 $\mu\text{g ml}^{-1}$ ICRF-193 (●), or nocodazole, ICRF-193 and 2 mM caffeine (■), nocodazole and 60 μM etoposide (○), or nocodazole, etoposide and 2 mM caffeine (△). *d*, Caffeine fails to affect G2 delay imposed by DNA damage in DM87 cells; mitotic accumulation of cells given 0.04 $\mu\text{g ml}^{-1}$ nocodazole (□), nocodazole immediately after 400 cGy ¹³⁷Cs γ -radiation (+), or nocodazole, 400 cGy ¹³⁷Cs γ -radiation and 2 mM caffeine (*). Mitotic accumulation (unsynchronized) was as for Fig. 1.

topoisomerase II inhibitors impose a checkpoint-dependent G2 block. After short times, the pseudomitotic figures (derived from cells near prophase at the time of topoisomerase II inhibition) are fairly well condensed; with increasing periods of accumulation, less condensed material is seen, derived from cells that were further from prophase when their topoisomerases were inhibited (Fig. 1a-c). This time-dependence of condensability indicates a progressive decatenatory activity in the late cycle.

This conclusion is only valid if ICRF-193 *in vivo* acts only as a topoisomerase inhibitor and lives up to its *in vitro* reputation of producing no DNA damage. Several independent pieces of evidence indicate that it does. First, the very sensitive alkaline unwinding assay reveals no significant DNA breakage in cells given ICRF-193, whereas the topoisomerase II inhibitor etoposide, which also inflicts DNA damage⁶, generates DNA strand breaks (Fig. 2a). (Neither were any DNA breaks detected previously in ICRF-193-treated cells using a less sensitive assay¹⁷.) Second, ICRF-193 cannot prevent condensation being attempted in G2 HeLa cells fused with mitotic cells (Fig. 2b), whereas etoposide prevents such premature condensation after G2/mitotic fusions unless caffeine is present. Thus, ICRF-193 appears to produce no damage response in this cellular assay. Third, the G2 chromosomes that partly condense in the presence of ICRF-193 and caffeine are intact (Figs 1a, b and 2c); whereas caffeine and etoposide combine to generate chromosomes with obvious breaks and exchanges (Fig. 2d). Fourth, cells arrested

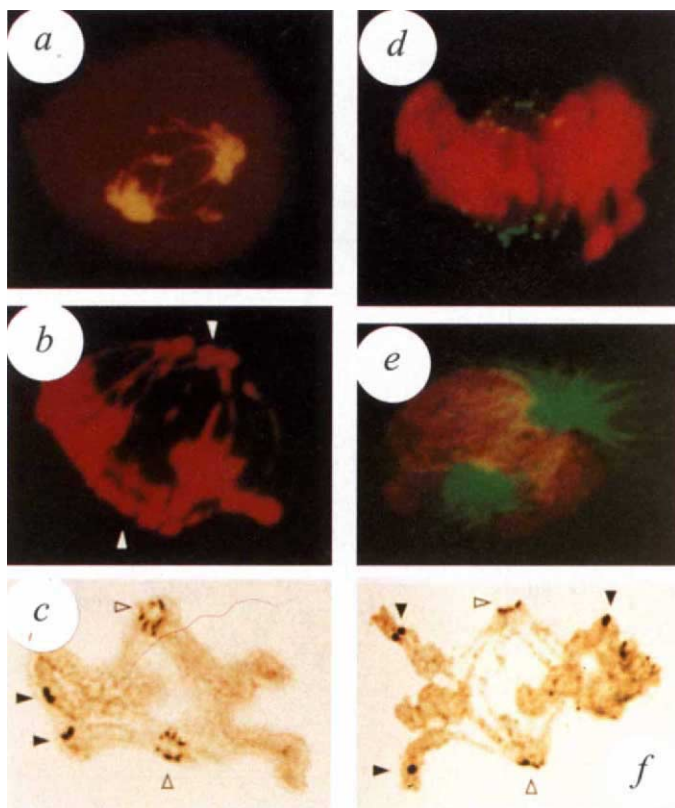


FIG. 4 Pseudo-anaphase mammalian cells with inadequately decatenated sister chromatids. *a*, Fluorescence microscopy, propidium iodide stain (chromatin, yellow; cytoplasm, red). *b*, Ribonuclease digestion, confocal microscopy, propidium iodide stain (chromatin, red). *d*, Confocal microscopy, cumulative image from series of 1 μm slices, kinetochores stained with CREST serum and FITC-conjugated goat anti-human IgG (green/yellow), propidium iodide counter-stain for chromatin (red). *e*, Same imaging, stained with anti-tubulin and FITC-conjugated goat anti-rat IgG, propidium iodide counterstain for chromatin. *c*, *f*, Bright-field microscopy, kinetochores visualized by silver impregnation³⁰. White triangles indicate apparent spindle orientation; black triangles, nucleolus organizer regions. Various techniques thus demonstrate that anaphase is attempted when chromatin is extensively catenated; mitotic spindles form and kinetochores migrate poleward. Anti-tubulin staining (*e*) reveals spindle formation; in agreement, propidium iodide staining (*a*, *b*) shows that chromatin is subjected to pulling forces. Silver impregnation/CREST antibody staining show that kinetochores are often coincident with chromatin separated from the chromatin bulk (*c*, *d*, *f*); kinetochores are also seen surrounding spindle poles (*c*). SVM muntjac cells (*a*, *b*, *c*, *f*) were accumulated in mitosis for 4 h with 2 $\mu\text{g ml}^{-1}$ ICRF-193, 2 mM caffeine, 0.15 $\mu\text{g ml}^{-1}$ nocodazole. Mitotic cells, washed with PBS, were allowed to attempt anaphase in medium containing ICRF-193 only, and fixed after 30 min. Muntjac cells were also washed with medium containing ICRF-193 and nocodazole and incubated in this medium for 2 h before nocodazole was removed; similar pseudo-anaphase figures were seen. PtK2 rat kangaroo cells were treated as above (*d*), or allowed to proceed into unarrested mitosis with caffeine and ICRF-193 (*e*). Cells were fixed with -20°C methanol for antibody staining *in situ*, or otherwise fixed and spread as for Fig. 1.

in G2 by ICRF-193 rapidly progress into mitosis with normal chromosomes once the block is removed, whereas cells treated with etoposide do not (Fig. 3*a*, *b*). Last, and most crucially, in Indian muntjac DM87 cells caffeine cannot, unusually, override cycle blocks imposed by replication inhibitors or ultraviolet radiation¹⁹ (although the cells retain other responses to caffeine²⁰). Similarly, the G2 block induced by γ -irradiation or by etoposide is caffeine-insensitive in DM87, but the block induced by ICRF-193 is not (Fig. 3*c*, *d*). Hence the ICRF-193-induced block is modulated through a system different from the damage-induced block.

We therefore propose that there are distinct checkpoint systems in mammalian G2, one sensitive to DNA damage and the other to the decatenation state of DNA. The catenation-sensitive system is leaky: this may be an example of the 'decision-by-committee' model for mitotic entry²¹, whereby a vote against progress (here, the catenation-sensitive checkpoint) may be overcome by an increased vote for progress from another part of the control system. It is uncertain whether a similar system exists in yeast, where other checkpoints are better characterized; temperature-sensitive *top2* mutants of *Schizosaccharomyces pombe* enter an abortive mitosis with very elongated chromosomes³, like those found here, without the benefit of checkpoint-evading agents or mutations. Whether there is a lag before such condensation is not reported.

Once cells have passed prophase, the decatenation-sensing checkpoint no longer operates. This is implied by the failure of ICRF-193 to prevent mitotic fusion partners inducing condensation (Fig. 2*b*), and by the inability of ICRF-193 or etoposide to prevent mitotic cells attempting cytokinesis^{16,22}. More strikingly, cells driven into mitosis without adequate decatenation, by ICRF-193 and caffeine, attempt to proceed (Fig. 4). Despite the bizarre abortive chromosomes, such cells are able to form mitotic spindles, and attempt anaphase separation of their catenated chromatids, with normal kinetochore positioning and separation. PtK1 rat kangaroo cells are also reported to attempt anaphase with imperfectly decatenated chromosomes²³.

We conclude that mammalian cells possess a previously unsuspected G2 checkpoint system, sensitive to the catenation of DNA or the decatenatory activity of topoisomerase II, either directly or through detection of some consequent change in chromatin preliminary to mitotic condensation, which ceases to operate after the G2/prophase transition. As some residual concatenation of sister chromatids must remain until the start of anaphase (as evidenced by the block to chromatid segregation imposed by inhibitors of topoisomerase II; refs 16, 22, 24), the checkpoint must sense the adequacy, not the completion, of decatenation. We propose that a main function of G2 is to allow decatenation of replicated DNA to become adequate. □

Received 20 June; accepted 19 October 1994.

1. Cook, P. R. *Cell* **66**, 627–635 (1991).
2. Ma, X. J., Saitoh, N. & Curtis, P. J. *J. biol. Chem.* **268**, 6182–6188 (1993).
3. Uemura, T. et al. *Cell* **50**, 917–925 (1987).
4. Wood, E. R. & Earnshaw, W. C. *J. Cell Biol.* **111**, 2839–2850 (1990).
5. Adachi, Y., Luke, M. & Laemmli, U. K. *Cell* **64**, 137–148 (1991).
6. Hirano, T. & Mitchison, T. J. *J. Cell Biol.* **120**, 601–612 (1993).
7. Kalwinsky, D. K., Look, A. T., Ducore, J. & Fridland, A. *Cancer Res.* **43**, 1592–1597 (1983).
8. Lock, R. B. & Ross, W. E. *Cancer Res.* **50**, 3761–3766 (1990).
9. Roca, J., Ishida, R., Berger, J. M., Andoh, T. & Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1781–1785 (1994).
10. Tolmach, L. J. & Busse, P. M. *Radiat. Res.* **82**, 374–392 (1980).
11. Yamashita, K. et al. *EMBO J.* **9**, 4331–4338 (1990).
12. Steinmann, K. E., Belinsky, G. S., Lee, D. & Schlegel, R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6843–6847 (1991).
13. Creighton, A. M., Hellman, K. & Whitecross, S. *Nature* **222**, 384–385 (1969).
14. Sharpe, H. B. A., Field, E. O. & Hellman, K. *Nature* **226**, 524–526 (1970).
15. Creighton, A. M. *Adv. Med. Oncol. Res. Educ.* **5**, 83–91 (1979).
16. Clarke, D. J., Johnson, R. T. & Downes, C. S. *J. Cell Sci.* **105**, 563–569 (1993).
17. Ishida, R. et al. *Cancer Res.* **51**, 4909–4916 (1991).
18. Ishimi, Y., Ishida, R. & Andoh, T. *Molec. cell. Biol.* **12**, 4007–4014 (1992).
19. Musk, S. R. R., Downes, C. S. & Johnson, R. T. *J. Cell Sci.* **90**, 591–599 (1988).
20. Musk, S. R. R., Pillidge, L., Johnson, R. T. & Downes, C. S. *Biochim. biophys. Acta* **1052**, 53–62 (1990).
21. Weinert, T. & Lydall, D. *Sem. Canc. Biol.* **4**, 129–140 (1993).
22. Downes, C. S., Mulliger, A. M. & Johnson, R. T. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2616–2620 (1991).
23. Gorbisky, G. J. *Cancer Res.* **54**, 1042–1048 (1994).
24. Shamu, C. E. & Murray, A. W. *J. Cell Biol.* **117**, 921–934 (1992).
25. Morgan, T. L. et al. *Expl. Cell Res.* **197**, 125–136 (1991).
26. Johnson, R. T., Downes, C. S. & Meyn, R. E. in *The Cell Cycle: A Practical Approach* (eds Fantes, P. & Brooks, R.) 1–24 (IRL, Oxford, 1994).
27. Ahnstrom, G. & Edvardsson, K.-A. *Int. J. Radiat. Biol.* **26**, 493–497 (1974).
28. Squires, S., Johnson, R. T. & Collins, A. R. S. *Mut. Res.* **95**, 389–404 (1982).
29. Johnson, R. T. & Rao, P. N. *Nature* **226**, 717–722 (1970).
30. Rufas, J. S., Gimenez-Abian, J., Suja, J. A. & Garcia de la Vega, C. *Genome* **29**, 706–712 (1987).

ACKNOWLEDGEMENTS. This work was supported by the CRC (R.T.J.), the British Technology Group (A.M.C.), the Commission of the European Community (J.F.G.A.) and the MRC (D.J.C.). We thank J. McCormick for MSU1.1 cells, J. Bedford for CREST antibody and B. Amos for anti-tubulin antibody.

To study the function of 77K subunit, all three CstF subunits were expressed from recombinant baculoviruses in various combinations. Partial or complete CstF complexes were purified by immunoaffinity chromatography using protein G Sepharose to which anti-50K monoclonal antibody (α 50K-antibody) (α 50K-PGS, Fig. 2a, lanes 1–3) or α 64K-antibody (α 64K-PGS, lanes 4–6)⁸ had been crosslinked. When Sf21 cells were coinfecting with recombinant 50K and 64K viruses, only the target proteins could be purified (lanes 1 and 4). In contrast, when the cells were coinfecting with recombinant 77K virus and either 50K (lane 2) or 64K virus (lane 5), both the target protein and 77K were copurified. These results suggest that the 77K subunit associates with both 50K and 64K subunits, thereby bridging these two subunits, and that the 50K and 64K subunits do not interact directly. Consistent with this, hetero-

trimeric CstF was formed when Sf21 cells were coinfecting with all three recombinant viruses (Fig. 2a, lanes 3 and 6). The functional activities of these partial or complete CstF complexes were then assayed in reconstituted *in vitro* cleavage reactions⁴. Only the heterotrimeric CstF complex (Fig. 2b, lanes 8 and 11) was able to substitute for human CstF (lane 4), with the level of activity correlated with the amount of the 64K subunit present in the CstF complexes (compare Fig. 2a, lanes 3 and 6; Fig. 2b, lanes 8 and 11). These results indicate that all three CstF subunits are essential for activity, and suggest that 77K function involves protein-protein interactions that bridge the other two subunits.

To confirm the role of the 77K subunit suggested by the above experiments, we examined interactions between CstF subunits *in vitro* (Fig. 3a). When mRNAs encoding flu-epitope tagged

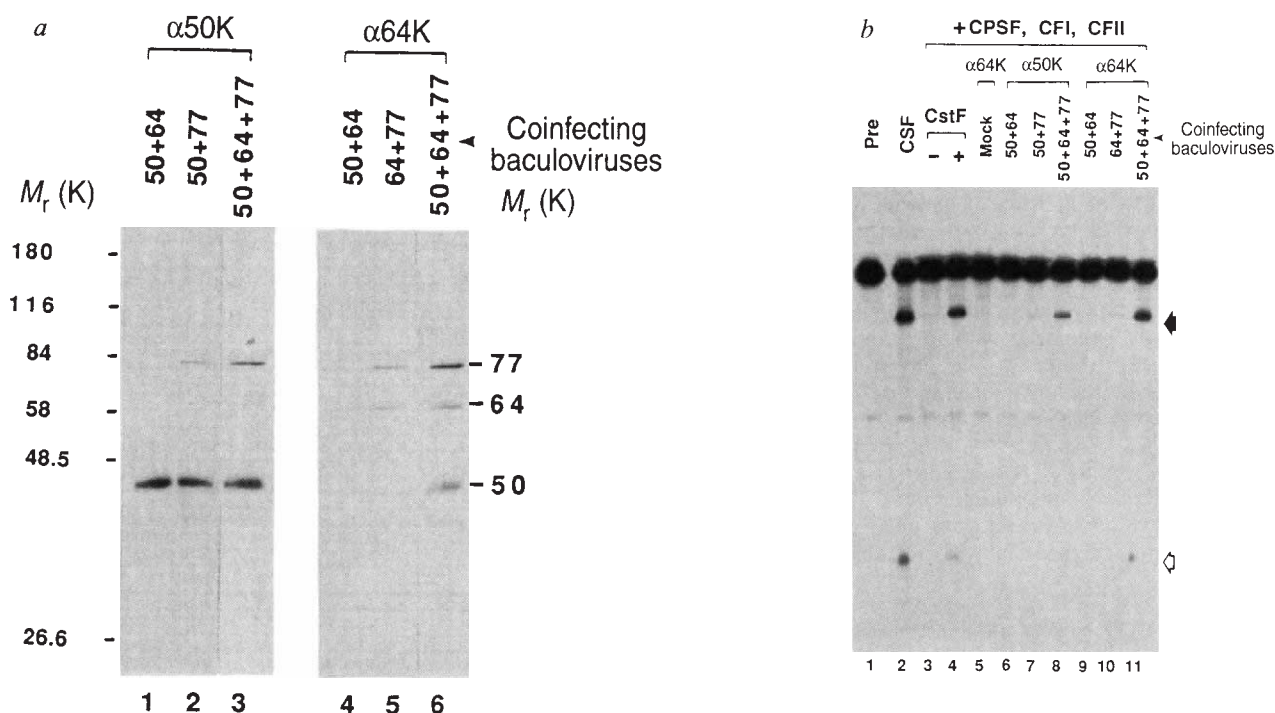


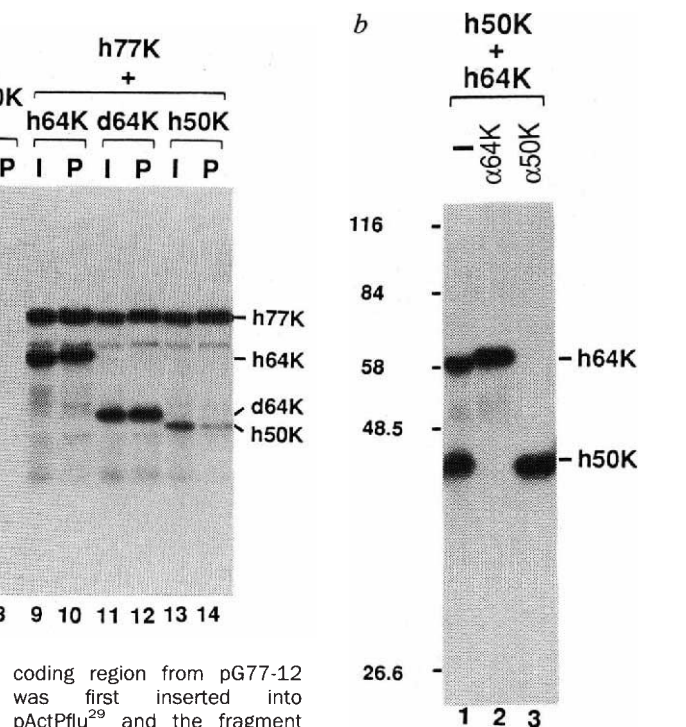
FIG. 2 Reconstitution of CstF complexes. a, Protein profiles of immunoaffinity purified CstF complexes. Sf21 cells were coinfecting with combinations of recombinant baculoviruses as indicated on top, and CstF complexes were purified by immunoaffinity chromatography using α 50K-PGS (lanes 1–3) or α 64K-PGS (lanes 4–6). CstF complexes were fractionated on a 10% SDS-polyacrylamide gel and stained with silver. Positions of the expressed CstF subunits (right) and protein size markers (left) are indicated. Identities of the purified proteins were confirmed by western blotting (data not shown). b, Functional analysis of the purified CstF complexes. CstF activity was assayed in the presence of the three other factors (CPSF, CFI and CFII) also required for cleavage of SV40 late pre-mRNA⁴. The reaction mixtures were supplemented with CstF complexes purified by α 50K-PGS (lanes 6–8) or α 64K-PGS (lanes 9–11). Cleavage specificity factor (CSF²⁴, lane 2) and partially purified CstF (lane 4) were used as positive controls and dialysis buffer (lane 3) and proteins purified from mock-infected Sf21 cells by α 64K-PGS (lane 5) were used as negative controls. Positions of SV40 late pre-mRNA (lane 1, Pre, 233 nt) and upstream (black arrow, 179 nt) and downstream (white arrow, 54 nt) cleavage products are indicated.

METHODS. Recombinant baculoviruses encoding each of the three CstF subunits were prepared by cotransfecting Sf21 insect cells with a mixture of 0.1 μ g of BaculoGold linearized virus DNA (PharMingen) and 3 μ g of transfer vector pEV55²⁵ containing cDNA sequence derived from pZ50–19¹¹, pZ64–18¹⁰ or pDS77 according to manufacturer's instructions. After cotransfection, recombinant baculoviruses were isolated and amplified as described²⁵. To express CstF subunit proteins, Sf21 cells were infected with mixtures of recombinant viruses at a multiplicity

of infection (MOI) of 5 for each virus. Forty hours after infection, cells were washed 3 times with ice-cold PBS and collected in 5 ml lysis buffer (50 mM Tris-HCl (pH 7.9), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, 1% NP40). After standing on ice for 30 min, cell debris was removed by spinning at 4,000 r.p.m. and 4 °C for 5 min. One ninth volume of glycerol was added to the supernatant, and the mixture was loaded on an appropriate monoclonal antibody (mAb)-PGS conjugate (0.1 ml) prepared by crosslinking mAb to PGS²⁷ and packed in a 1 ml pipette tip. After extensive washing with both lysis buffer and wash buffer (same as lysis buffer but containing 300 mM NaCl and no NP40), bound proteins were eluted with 0.5 ml of elution buffer (50 mM Tris-HCl (pH 8.5), 1.5 M NaCl, 1 mM EDTA, 1 mM DTT, 50% (v/v) ethylene glycol) and dialysed against buffer D (20 mM HEPES-NaOH (pH 7.9), 50 mM (NH₄)₂SO₄, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 20% (v/v) glycerol). The molar ratios between the three subunit proteins copurified by α 50K-PGS and by α 64K-PGS were somewhat different, although the ratio was almost unity when the complex was purified by α 64K-PGS. The disproportionate recovery of the three subunits in the α 50K-PGS purified preparation may be because the 50K subunit was more efficiently expressed and/or more stable than the other two subunits in Sf21 cells. To analyse protein profiles, 10 μ l of each protein complex was fractionated on a 10% SDS-polyacrylamide gel and stained with silver. To assay CstF activity, 2 μ l of purified CstF complex was incubated with 2 μ l each of CPSF- and CFI+CFII-containing fractions, which had been obtained by Mono Q chromatography⁴, and 3 ng of ³²P-labelled SV40 late pre-mRNA⁸. Reaction products were fractionated on a 8.3 M urea–5% polyacrylamide gel.

FIG. 3 Formation of CstF complex *in vitro*. **a**, The 77K subunit binds the other two CstF subunits *in vitro*. mRNAs encoding flu-tagged h77K (lanes 1 and 2), h64K (lanes 3 and 4), d64K (lanes 5 and 6) or h50K (lanes 7 and 8) alone, or a mixture of h77K mRNA and h64K (lanes 9 and 10), d64K (lanes 11 and 12) or h50K mRNA (lanes 13 and 14) were translated *in vitro*. The input proteins (lanes I) or proteins immunoprecipitated with α flu-mAb (lanes P) were fractionated on a 10% SDS-polyacrylamide gel. Positions of protein size markers and translated CstF subunits are indicated on the left and right, respectively. **b**, The 64K and 50K subunits do not interact *in vitro*. A mixture of h50K and h64K mRNAs was translated *in vitro*. The input proteins (lane I) or proteins immunoprecipitated with α 64K- (lane α 64K) or α 50K-mAb (lane α 50K) were fractionated on a 10% SDS-polyacrylamide gel.

METHODS. For efficient translation of mRNAs *in vitro*, 5'-untranslated regions of the cDNAs encoding h77K, h64K and h50K were replaced by that of β -globin gene as described²⁸. For the expression of h50K, pZ50-19¹¹ cDNA sequence was first amplified by PCR using 0.53 μ g 5'-terminal oligonucleotide (GL50: AAACAGACACCATGGACAGAACC-AAAGTGGGC) and 0.33 μ g T7 promoter primer. The PCR product was subcloned into pGEM-3 (pGL-h50K). For the expression of h64K, pZ64-18¹⁰ cDNA sequence was amplified by PCR using 0.48 μ g 5'-terminal oligonucleotide (GL64: AAACAGACACCATGGCGGGTTTGACTGTG) and 0.30 μ g T3 promoter primer. The PCR product was subcloned as above (pGL-h64K). For the expression of flu-tagged²⁶ h77K, the



coding region from pG77-12 was first inserted into pActPflu²⁹ and the fragment encoding the flu epitope-h77K fusion transferred to pGEM-3 (pGL-h77K). Plasmids were linearized and transcribed by SP6 RNA polymerase (NEB) to produce mRNAs for *in vitro* translation. For expression of d64K, a *Drosophila* 64K cDNA (pZd64-19) was isolated from a *Drosophila* ovary cDNA library in λ ZAP II (Y.T. and J.L.M., manuscript in preparation). To synthesize d64K mRNA, pZd64-19 was linearized and transcribed by T3 RNA polymerase (Promega). mRNAs were translated by rabbit reticulocyte lysate (Promega) in the presence of 35 S-Met (NEN). Immunoprecipitation experiments were as described previously⁸.

human 77K (flu-h77K) plus human (h64K, lanes 9 and 10) or *Drosophila* (d64K, lanes 11 and 12) 64K subunit or human 50K subunit (h50K, lanes 13 and 14) were cotranslated and the products immunoprecipitated with anti-flu epitope monoclonal antibody (α flu-antibody, lanes 10, 12 and 14), each of the three proteins was efficiently coprecipitated with h77K. In contrast, when mRNAs encoding h50K and either h64K (Fig. 3b, lane 1) or d64K (not shown) were cotranslated and the products immunoprecipitated with α 64K- (lane 2) or α 50K-antibody (lane 3), only the target protein was precipitated. These results confirm those obtained *in vivo*, and indicate that h77K associates with both 50K and 64K subunits. That h77K also interacts with d64K suggests that the *Drosophila* homologue of the 77K subunit (see below) functions similarly.

The primary structure of the 77K subunit predicted from the cDNA sequence (Fig. 1) exhibits some characteristic features. First, although the protein contains two hydrophobic regions (amino acids 554-575 and 605-636) and a Pro-rich domain (579-632, underlined in Fig. 1), 77K is hydrophilic overall, as residues with polar side chains account for 52.7% of the total. Clusters of acidic (689-696), basic (77-121, 383-408 and 709-717) and charged (4-33, 184-232 and 503-553) residues are also found, particularly in the amino terminus of the protein. In addition, a putative bipartite nuclear localization signal (NLS)¹⁵ is present at 386-402, suggesting that because the other two subunits of CstF do not have NLS-like sequences^{10,11}, despite their nuclear localization⁶, formation of heterotrimeric CstF in the cytoplasm may be a prerequisite for its transport into the nucleus.

A search of the genbank database (June 1994) revealed that 77K shares extensive homology with the protein encoded by the *Drosophila* *su(f)* gene (733 amino acids, M_r 84K; identity

56.2%, similarity 69.4%) and moderate homology with the *S. cerevisiae* RNA14 protein (636 amino acids, M_r 75,295; 24.3%, 37.2%) (Fig. 4). Based on the extent of identity, 77K can be divided into five domains: N terminus (N, amino acids 1-76), highly conserved domains I (I, 77-306) and II (II, 371-597), a less conserved region (M, 307-370) and the carboxy terminus, which is lacking in RNA14 (C, 598-717). Of the 15 Pro residues present in the Pro-rich domain, which encompasses the junction between domains II and C, 14 are conserved between the 77K subunit and *su(f)* protein (Fig. 4, asterisks), suggesting the importance of these residues for function.

The significance of the 77K-RNA14 homology is strengthened by the fact that both human and *Drosophila* 64K subunits share significant homology with the protein encoded by the *S. cerevisiae* RNA15 gene, which has been linked genetically to RNA14¹⁴. Although RNA15 (296 amino acids) is noticeably smaller than human (577 amino acids)¹⁰ and *Drosophila* (419 amino acids; Y.T. and J.L.M., manuscript in preparation) 64K proteins, the RNP-type RNA-binding domains (RBDs) present at the N termini of the proteins are highly conserved between both h64K and RNA15 (identity 42.5%, similarity 62.5%) and d64K and RNA15 (43.8%, 67.5%). Indeed, the RNA15 RBD is more closely related to the h64K and d64K RBDs than any of the greater than 100 other known RBDs. These results strongly suggest that RNA14 and RNA15 correspond to the 77K/*su(f)* and 64K subunits of human and *Drosophila* CstF, respectively, and provide evidence that the mechanism of 3'-pre-mRNA processing is conserved from yeast through *Drosophila* to man.

It has long been known that mutations in certain *Drosophila* genes, called modifier genes, modulate the phenotypes of mutations in unlinked genes caused by insertion of transposable

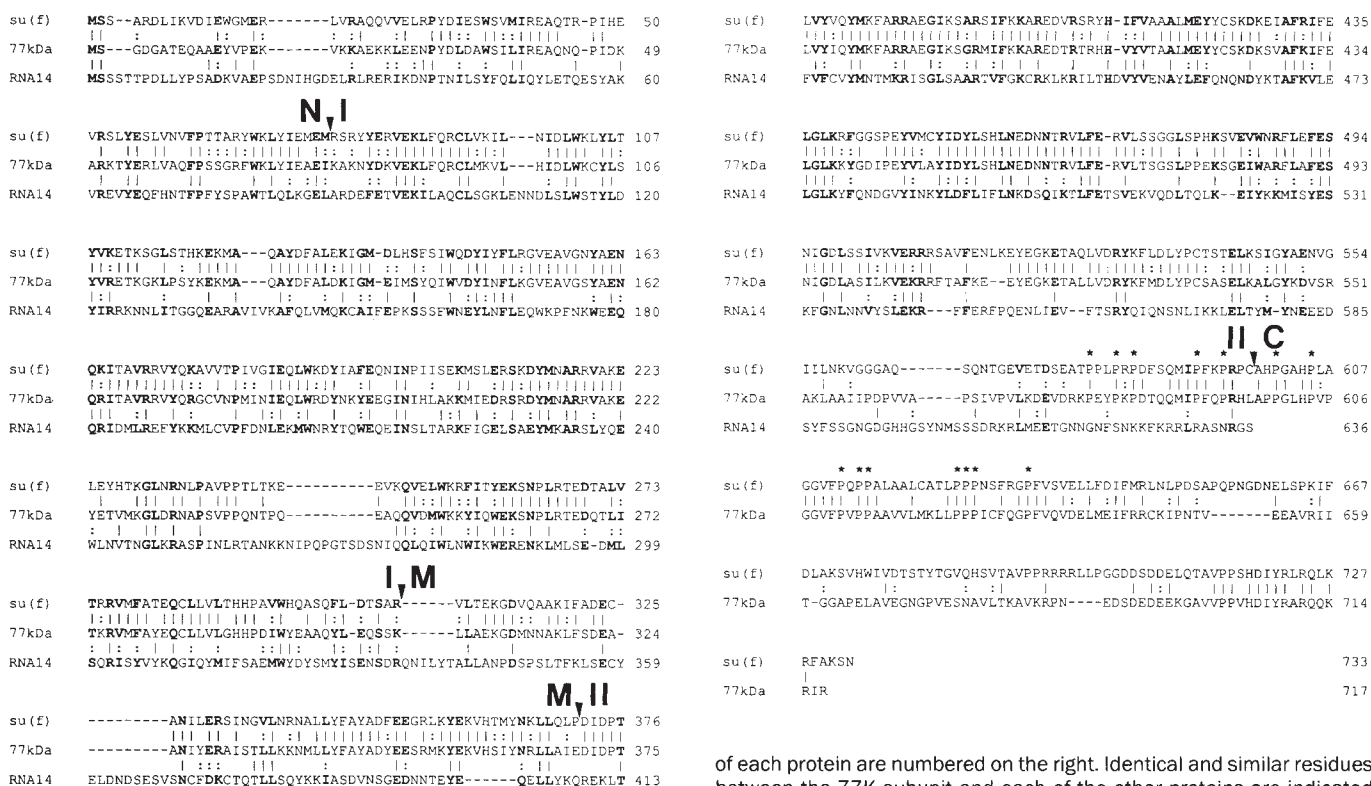


FIG. 4 The 77K subunit shares homology with *Drosophila su(f)* and yeast RNA14 proteins. Amino-acid sequence of the 77K subunit is optimally aligned with those of the *Drosophila su(f)* protein¹² (top) and yeast RNA14 protein¹⁴ (bottom) using the fasta program³⁰, and amino acids

elements into introns^{13,16}. The long terminal repeats (LTRs) of transposable elements, such as gypsy and copia, contain polyadenylation signals¹⁶, and it has been shown that insertion of gypsy into an intron of the *hsp82* gene results in truncation of transcription at the 5'-LTR¹⁷. Mutations in *su(f)* can suppress this truncated transcription to increase wild-type mRNA levels, suggesting that modulation of an RNA processing event such as polyadenylation, or perhaps splicing (for example, ref. 18), plays an important role in modifying gene expression. Note that besides the modifier functions exerted by some mutations in *su(f)*¹³, other mutations cause temperature-sensitive lethality^{16,18}, which can be associated with female sterility²⁰, and null mutations cause cell autonomous, unconditional lethality^{12,21}. Interestingly, a single allele of *su(f)* not only can suppress mutations caused by insertion of the gypsy element into any of several genes, but also can enhance a mutation caused by insertion of the copia element into the *white* gene¹³. These observations strongly suggest that *su(f)* plays a vital role in modulating gene expression, and the results described here indicate that this occurs at the level of polyadenylation. These properties of *su(f)* also suggest that small changes in activity can have significant and gene specific effects on polyadenylation efficiency and, consequently, mRNA production. Our findings here that 77K subunit function involves interactions with at least two different proteins, and previous findings that CstF and CPSF bind pre-mRNA in a highly cooperative and AAUAAA-dependent manner^{9,22}, probably provide at least part of the explanation. Together, these results indicate that CstF may be able to regulate expression of specific genes. By analogy with *su(f)*, fluctuations in the amounts or activity of CstF subunits may well change polyadenylation efficiencies of different genes, perhaps in some cases resulting in differential usage of alternative poly(A) sites. □

of each protein are numbered on the right. Identical and similar residues between the 77K subunit and each of the other proteins are indicated by lines and dots, respectively. Identical and similar residues present in all three proteins are shown in bold. Boundaries which separate the N terminus (N), highly conserved domains I (I) and II (II), the less conserved region between them (M) and the C-terminus (C) are indicated by arrowheads. Pro residues in the Pro-rich domain that are conserved between 77K and *su(f)* proteins are indicated by asterisks.

Received 30 August; accepted 14 October 1994.

- Wahle, E. & Keller, W. A. *Rev. Biochem.* **61**, 419–440 (1992).
- Manley, J. L. & Proudfoot, N. J. *Genes Dev.* **8**, 259–264 (1994).
- Christofori, G. & Keller, W. *Cell* **54**, 875–889 (1988).
- Takagaki, Y., Ryner, L. C. & Manley, J. L. *Genes Dev.* **3**, 1711–1724 (1989).
- Gilmartin, G. M. & Nevins, J. R. *Genes Dev.* **3**, 2180–2189 (1989).
- Raabe, T., Bollum, F. J. & Manley, J. L. *Nature* **353**, 229–234 (1991).
- Wahle, E., Martin, G., Schiltz, E. & Keller, W. *EMBO J.* **10**, 4251–4257 (1991).
- Takagaki, Y., Manley, J. L., MacDonald, C. C., Wilusz, J. & Shenk, T. *Genes Dev.* **4**, 2112–2120 (1990).
- Gilmartin, G. M. & Nevins, J. R. *Molec. cell. Biol.* **11**, 2432–2438 (1991).
- Takagaki, Y., MacDonald, C. C., Shenk, T. & Manley, J. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1403–1407 (1992).
- Takagaki, Y. & Manley, J. L. *J. biol. Chem.* **267**, 23471–23474 (1992).
- Mitchelson, A., Simonelig, M., Williams, C. & O'Hare, K. *Genes Dev.* **7**, 241–249 (1993).
- Rutledge, B. J., Mortin, M. A., Schwarz, E., Thierry-Mieg, D. & Meselson, M. *Genetics* **119**, 391–397 (1988).
- Minvielle-Sebastia, L., Winsor, B., Bonneaud, N. & Lacroute, F. *Molec. cell. Biol.* **11**, 3075–3087 (1991).
- Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. *Cell* **64**, 615–623 (1991).
- Lindsley, D. L. & Zimm, G. *The Genome of Drosophila Melanogaster* (Academic, San Diego, 1992).
- Dorsett, D., Viglianti, G. A., Rutledge, B. J. & Meselson, M. *Genes Dev.* **3**, 454–468 (1989).
- Zachar, J., Chou, T.-B. & Bingham, P. M. *EMBO J.* **6**, 4105–4111 (1987).
- Dudick, M. E., Wright, T. R. F. & Brothers, L. L. *Genetics* **76**, 487–510 (1974).
- Wilson, T. G. *J. Embryol. exp. Morph.* **55**, 247–256 (1980).
- Driver, A., Lacey, S. F., Cullingford, T. E., Mitchelson, A. & O'Hare, K. *Molec. Gen. Genet.* **220**, 49–52 (1989).
- Murthy, K. G. K. & Manley, J. L. *J. biol. Chem.* **267**, 14804–14811 (1992).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Takagaki, Y., Ryner, L. C. & Manley, J. L. *Cell* **52**, 731–742 (1988).
- Miller, D. W., Safer, P. & Miller, L. K. in *Genetic Engineering* vol. 8 (eds Setlow, J. K. & Hollaender, A.) 277–298 (Plenum, New York, 1986).
- Field, J. et al. *Molec. cell. Biol.* **8**, 2159–2165 (1988).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
- Aurora, R. & Herr, W. *Molec. cell. Biol.* **12**, 455–467 (1992).
- Han, K. & Manley, J. L. *Genes Dev.* **7**, 491–503 (1993).
- Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444–2448 (1988).

ACKNOWLEDGEMENTS. We thank C. Prives and I. Reynisdottir for discussions on the baculovirus expression system and for providing reagents, S. M. Mount and R. Mancebo for sharing with us their knowledge of *Drosophila* genetics, S. M. Mount for comments on the manuscript, T. Hazlett for supplying a *Drosophila* ovary cDNA library, K. Han and J. Colgan for supplying pActPfl, Z. Lai and E. Freulich for technical assistance and M. Hosford for preparing the manuscript. This work was supported by a NIH grant.

Protein disaggregation mediated by heat-shock protein Hsp104

Dawn A. Parsell^{*†‡}, Anthony S. Kowal[†],
Mike A. Singer[§] & Susan Lindquist^{*†||}

Departments of ^{*} Molecular Genetics and Cell Biology, [§] Pathology, and [†] The Howard Hughes Medical Institute, The University of Chicago, MC1028 N339, 5841 S. Maryland Avenue, Chicago, Illinois 60637, USA

THE heat-inducible members of the Hsp100 (or Clp) family of proteins share a common function in helping organisms to survive extreme stress, but the basic mechanism through which these proteins function is not understood¹⁻⁵. Hsp104 protects cells against a variety of stresses, under many physiological conditions^{6,7}, and its function has been evolutionarily conserved, at least from *Saccharomyces cerevisiae* to *Arabidopsis thaliana*²⁵. Homology with the *Escherichia coli* ClpA protein suggests that Hsp104 may provide stress tolerance by helping to rid the cell of heat-denatured proteins through proteolysis¹. But genetic analysis indicates that Hsp104 may function like Hsp70 as a molecular chaperone⁸. Here we investigate the role of Hsp104 *in vivo* using a temperature-sensitive *Vibrio harveyi* luciferase-fusion protein as a test substrate⁹. We find that Hsp104 does not protect luciferase from thermal denaturation, nor does it promote proteolysis of luciferase. Rather, Hsp104 functions in a manner not previously described for other heat-shock proteins: it mediates the resolubilization of heat-inactivated luciferase from insoluble aggregates.

We first determined the conditions under which our test substrate would be inactivated while both wild-type and mutant cells remained viable. We found that pretreatment at 37 °C partially protected luciferase from inactivation at extreme temperatures (Fig. 1a) but that luciferase was always inactivated to a comparable extent in wild-type (*HSP104*) and mutant (*hsp104*) cells. We conclude that Hsp104 does not play a significant part in this protection.

To investigate the proposed role of Hsp104 in proteolysis of heat-damaged proteins, we analysed the stability of luciferase after a severe heat shock. Cells were pretreated at 37 °C for 30 min to induce synthesis of heat-shock proteins, shifted to 44 °C for 1 hour, then returned to 25 °C. Cycloheximide was added to block protein synthesis during the recovery period at 25 °C. Luciferase levels remained constant in both wild-type and *hsp104* mutant cells throughout the experiment (Fig. 1b). Thus, Hsp104 does not appear to affect turnover of the heat-inactivated test substrate.

We next investigated whether Hsp104 functions in reactivating heat-damaged proteins. Cells were pretreated at 37 °C, heat-shocked at 44 °C, and then allowed to recover at 25 °C in the absence of new protein synthesis. Wild-type cells recovered 90% of their initial luciferase activity, but *hsp104* mutant cells failed to reactivate the enzyme (Fig. 2a). This result suggests that, rather than serving to rid the cell of stress-damaged proteins, Hsp104 functions to promote their proper renaturation and reactivation. To test the importance of the two ATP-binding sites of Hsp104 for this function, we examined luciferase reactivation in strains expressing *hsp104* alleles containing single inactivating point mutations. As previously demonstrated for thermotolerance², the function of both ATP-binding sites was required for luciferase reactivation (Fig. 2b).

The failure of the *hsp104* strain to reactivate temperature-sensitive bacterial luciferase was reminiscent of the inability of

the *E. coli hsp70 (dnaK)* mutants to restore the activity of denatured firefly luciferase¹⁰. To investigate the role of yeast Hsp70 proteins in reactivation, we examined mutants in which Hsp70 levels were diminished either by reducing constitutive expression (*ssa1ssa2*) or eliminating heat-inducible expression (*ssa1ssa3ssa4*). The basal level of luciferase activity was reduced approximately fourfold in the *ssa1ssa2* strain. In both of the *hsp70* mutant strains, the enzyme was inactivated by heat shock at 44 °C to roughly the same extent as in the wild type. In contrast to the *hsp104* mutant, however, the *ssa1ssa3ssa4* mutant recovered luciferase activity at 25 °C as efficiently as the wild type. The *ssa1ssa2* mutant also recovered substantially (Fig. 2c). Only when Hsp104 was eliminated in the *hsp70* mutant background were cells unable to reactivate luciferase.

To investigate the state of the test substrate in wild-type, *hsp70*, and *hsp104* mutant cells, lysates were prepared at various times during the luciferase inactivation/reactivation experiment. These lysates were centrifuged at high speed and the fraction of luciferase remaining in the supernatant determined. Heat inactivation gave a dramatic decrease in the concentration of soluble luciferase in all cell types (Fig. 3). In wild-type and *hsp70* mutant

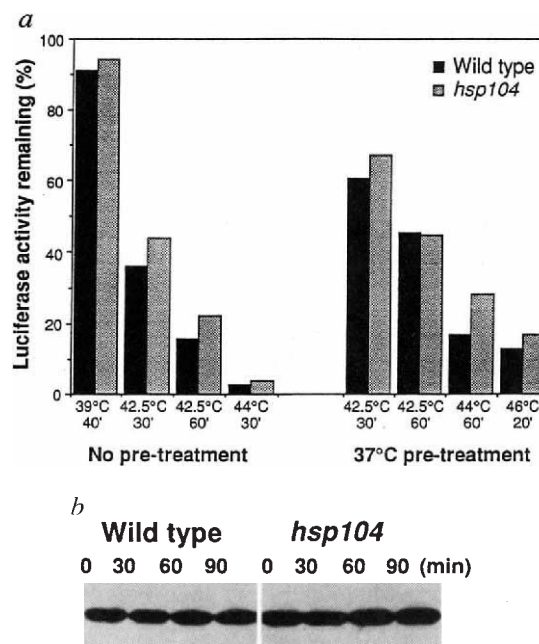


FIG. 1 **a**, Heat inactivation of luciferase in wild-type and *hsp104* mutant yeast cells. Wild-type (strain A1456) and *hsp104* mutant (strain A1458) cells, bearing 2 μ plasmids that direct the expression of a temperature-sensitive luciferase-fusion protein⁹ from the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (pGPD_{luc}(HIS)), were grown at 25 °C to mid-log phase in minimal dextrose medium¹⁹ supplemented with nutrients. Cultures were shifted to high temperature for the indicated lengths of time either with (right) or without (left) a preconditioning treatment of 37 °C for 30 min. Luciferase activity was assayed *in vivo* by adding 10 μ l of *n*-decylaldehyde (Sigma) to 1 ml log-phase cultures. Light emission was measured immediately with a Tropic luminometer. Activity levels are expressed as a percentage of the activity before the heat treatment. All strains are derivatives of W303 (gift from R. J. Rothstein). **b**, Luciferase protein levels do not change during heat inactivation and recovery. Wild-type and *hsp104* mutant cells bearing the luciferase expression plasmid were grown at 25 °C, pretreated at 37 °C for 30 min to induce Hsp expression, shifted to 44 °C for 60 min to inactivate luciferase and then allowed to recover at 25 °C. Cycloheximide (10 μ g ml⁻¹) was added during the last 10 min of the 44 °C treatment to block further protein synthesis. Total cellular proteins were extracted from cells at intervals during recovery using glass-bead lysis²⁰ and separated by SDS-PAGE on 10% gels²¹, transferred to Immobilon membranes (Millipore) and reacted with luciferase-specific antiserum. Immune complexes were visualized using horseradish-peroxidase-conjugated protein A and ECL detection system (Amersham).

[‡] Present address: Connective Therapeutics Inc., 3400 West Bayshore Road, Palo Alto, California 94303, USA.

^{||} To whom correspondence should be addressed.

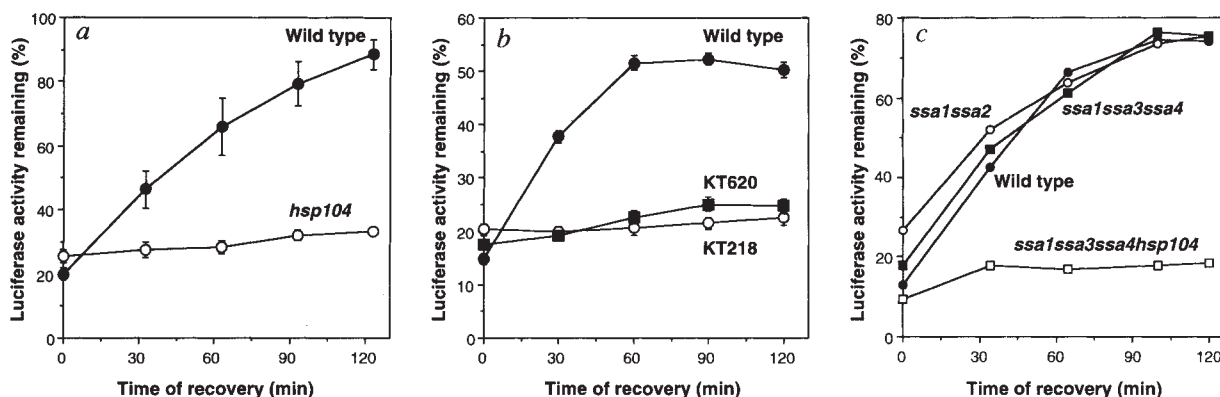


FIG. 2 Reactivation of heat-damaged luciferase requires Hsp104. **a**, Wild-type (filled circles) and *hsp104* mutant (circles) cells bearing the luciferase expression plasmid were treated as described for Fig. 1b and luciferase activity assayed *in vivo* as in Fig. 1a. Activity, expressed as a percentage of the luciferase activity in cells just before the 44 °C heat shock, was measured in duplicate for each of three separate cultures for each strain. **b**, *hsp104* mutant cells carrying the luciferase expression construct and a centromeric plasmid encoding either the wild-type HSP104 gene (filled circles; strain A1625) or an *hsp104* allele containing a mutation (lysine to threonine) which destroys the function of the

first (circles; strain A1626) or second (filled squares; strain A1627) ATP-binding site of the protein, were assayed for luciferase activity. Expression of all three alleles was driven by the HSP104 promoter. **c**, *S. cerevisiae* strains^{8,22} carrying mutations in the heat-inducible (filled squares; strain A1592: *ssa1ssa3ssa4*) or constitutive (circles; strain A1591: *ssa1ssa2*) cytosolic/nuclear *hsp70* genes were assayed for luciferase activity. For comparison, luciferase reactivation was measured in a wild-type strain (filled circles; strain A1630) and in a strain carrying mutation in both the heat-inducible Hsp70 genes and in Hsp104 (squares; A1593: *ssa1ssa3ssa4 hsp104*).

cells, luciferase was resolubilized during the course of recovery at 25 °C but there was little resolubilization of luciferase in *hsp104* mutant cells. Thus, it appears that Hsp104 promotes the restoration of luciferase activity by facilitating resolubilization of this protein from insoluble aggregates.

To determine whether the function of Hsp104 in promoting the resolubilization of luciferase reflects its general role in stress tolerance, cells were examined by electron microscopy. Wild-type and mutant cells that did not contain luciferase-expressing plasmids were subjected to the conditions used for the luciferase inactivation/reactivation experiments. The morphology of both wild-type and *hsp104* cells was normal at 25 °C (Fig. 4a). After heat treatment at 44 °C, damage was evidenced by the accumulation of large electron-dense aggregates in the cytoplasm and, more dramatically, in the nucleus. The severity of this damage was indistinguishable in the two cell types. In wild-type cells, this damage was resolved during recovery at 25 °C. After 120 min at 25 °C, damage had disappeared from almost all wild-type cells (Fig. 4a, b). Although *hsp70* strains continued to show abnormalities, particularly in membrane structures (A.S.K. and S.L.,

unpublished), after 120 min of recovery at 25 °C aggregation damage was resolved almost as well as in wild-type cells (Fig. 4c).

In contrast, *hsp104* mutant cells did not resolve the aggregation damage generated by heat treatment at 44 °C. Even after 120 min of recovery, most *hsp104* cells were still almost as damaged as they were immediately after heat shock (Fig. 4a, b). The aggregates seen here are not simply a consequence of cell death as both wild-type and *hsp104* mutant cells are viable under these conditions (data not shown). We believe, however, that these aggregates are directly related to the primary lethal lesion caused by treatment at higher temperatures, because similar aggregates are observed in wild-type cells at lethal temperatures¹¹.

In summary, we have shown that Hsp104 functions to reactivate a heat-denatured protein by promoting its resolubilization from an aggregated form. Ultrastructural analysis of heat-shocked cells indicates that this reactivation function is broad and represents the mechanism by which Hsp104 promotes survival under extreme stress. Our findings also help to distinguish the roles of Hsp70 and Hsp104 in thermotolerance. Strains lack-

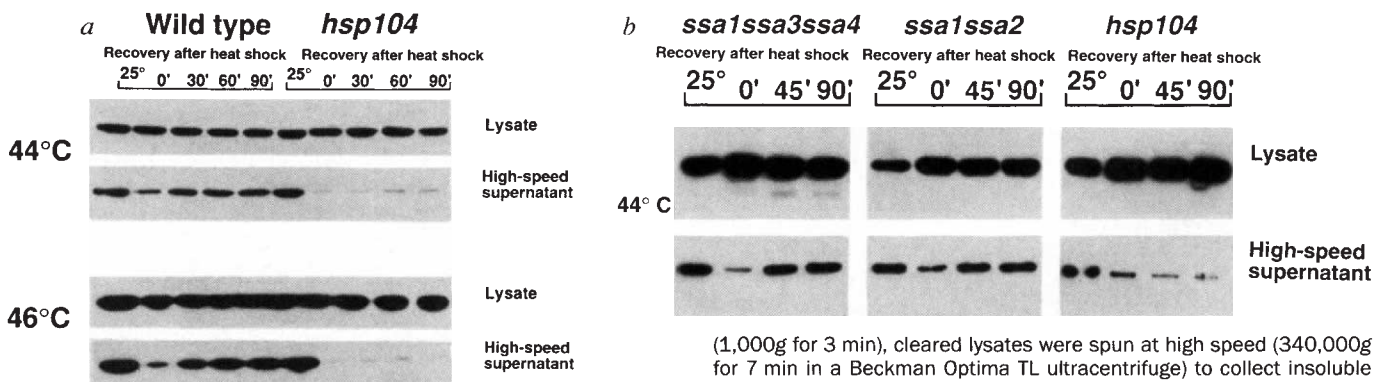
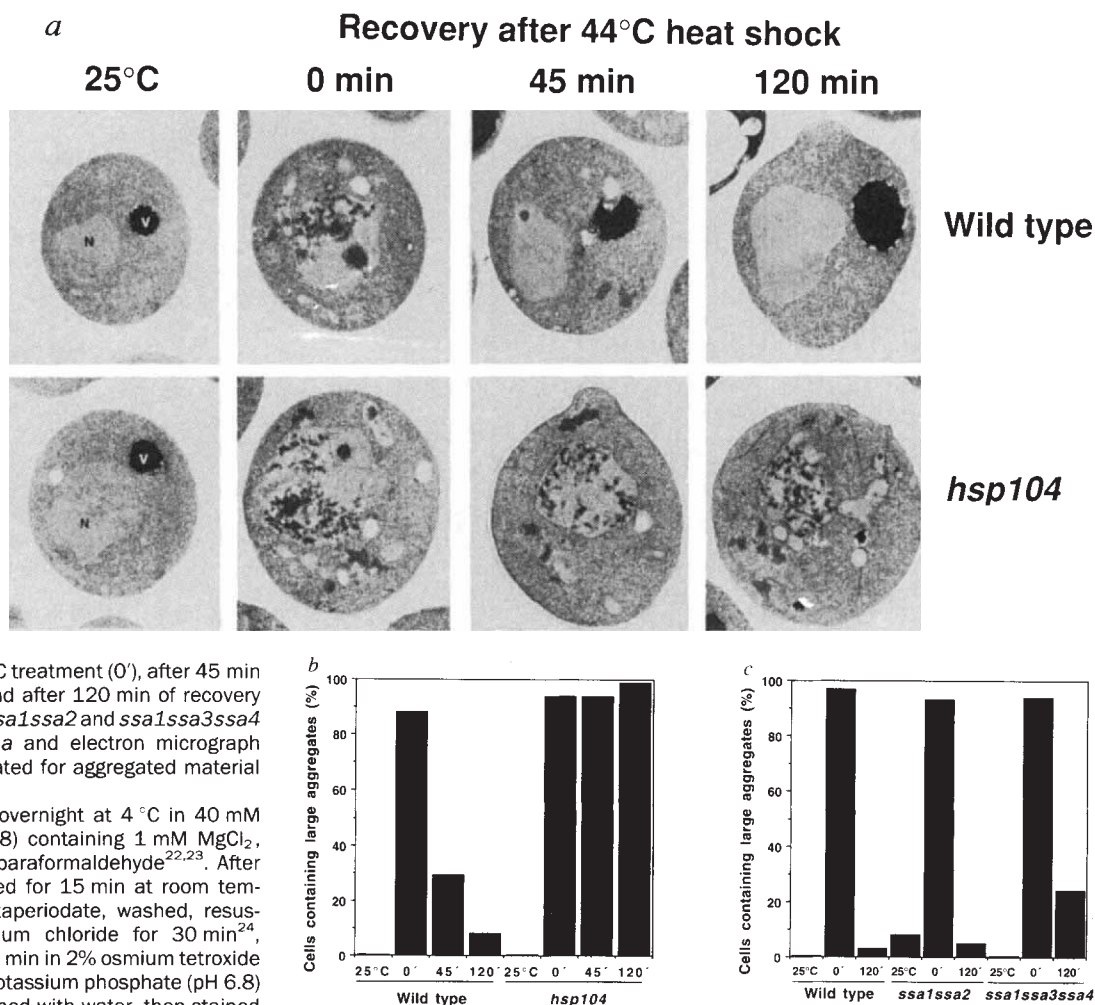


FIG. 3 Hsp104 promotes the resolubilization of aggregated luciferase. **a**, Wild-type and *hsp104* mutant cells carrying the luciferase expression plasmid were pretreated at 37 °C for 30 min, heat-shocked at either 44 °C for 60 min or at 46 °C for 24 min, then allowed to recover at 25 °C after adding cycloheximide and lysed as for Fig. 1 using glass-bead lysis. After removing unbroken cells by low-speed centrifugation

(1,000g for 3 min), cleared lysates were spun at high speed (340,000g for 7 min in a Beckman Optima TL ultracentrifuge) to collect insoluble aggregates. Total protein from the cleared lysates and high-speed supernatants was western blotted and probed as described in Fig. 1 legend. **b**, *ssa1ssa3ssa4* and *ssa1ssa2* mutants were treated as in **a**, with heat shock at 44 °C for 60 min. Coomassie staining of the blot demonstrated equal loading of all but the 25 °C samples, which were underloaded (data not shown). Because *ssa1ssa2* strains contained less luciferase, exposure was longer for these samples.

FIG. 4 *hsp104* mutants are unable to resolve heat-induced aggregates that are visible by electron microscopy. **a**, Wild-type and *hsp104* mutant cells were grown at 25 °C to mid-log phase in rich dextrose medium¹⁹. Cells were pretreated, heat-shocked at 44 °C for 60 min, and allowed to recover as in Fig. 3. Portions of the cultures were collected before any heat treatment (25 °C), immediately after the 44 °C heat shock (0 min), after 45 min of recovery at 25 °C (45 min) and after 120 min of recovery at 25 °C (120 min). **b**, For each of the samples in **a**, electron micrograph fields of 52–100 cells were screened (double blind) for aggregated material at 25 °C before heat treatment, immediately after 44 °C treatment (0'), after 45 min of recovery at 25 °C (45'), and after 120 min of recovery at 25 °C (120'). **c**, Wild-type, *ssa1ssa2* and *ssa1ssa3ssa4* cells were prepared as for **a** and electron micrograph fields of 31–151 cells evaluated for aggregated material as for **b**.

METHODS. Cells were fixed overnight at 4 °C in 40 mM potassium phosphate (pH 6.8) containing 1 mM MgCl₂, 1% glutaraldehyde and 1% paraformaldehyde^{22,23}. After washing, cells were incubated for 15 min at room temperature in 1% sodium metaperiodate, washed, resuspended in 50 mM ammonium chloride for 30 min²⁴, washed, and post-fixed for 60 min in 2% osmium tetroxide in buffer containing 40 mM potassium phosphate (pH 6.8) 1 mM MgCl₂. Cells were washed with water, then stained overnight at 4 °C with 0.5% uranyl acetate²² in 40 mM sodium malate (pH 5.2), washed, dehydrated with ethanol, and washed three times with 100% propylene oxide, then gradually infiltrated with resin.



Sections were stained with 4% aqueous uranyl acetate, washed, stained with 0.2% lead citrate, and viewed with a Phillips CM10 transmission electron microscope at 60 kV. Magnification, $\times 7,440$.

ing either the constitutive or the heat-inducible Hsp70 proteins recover luciferase activity almost to wild-type levels and resolve general aggregation damage nearly as well as wild-type cells. Although these results are surprising, given that *dnaK* mutants are unable to reactivate heat-denatured firefly luciferase in *E. coli*, they agree with the properties of DnaK *in vitro*¹⁰. DnaK can only reactivate firefly luciferase if luciferase has been prevented, by DnaJ, from aggregating. Under our conditions, heat-denatured proteins could presumably not be reactivated by Hsp70 because they were already aggregated (Figs 3, 4). A disaggregation activity has been reported for DnaK with an RNA polymerase substrate¹², but as core polymerase forms oligomers under such conditions¹³, this activity could reflect the ability of DnaK to disassemble certain types of protein oligomers^{13,15}. Although Hsp70 may promote the dissolution of some protein aggregates, we suggest that its primary function in thermotolerance is to help prevent such aggregates from forming^{10,12,16}. Hsp104, in contrast, has a remarkable capacity to rescue proteins from aggregates once they have formed. Hsp70 and Hsp104 can partially compensate for one another⁸ because they both affect the partitioning of damaged substrates between a denatured, soluble state and an insoluble, aggregated one. Thus, although their biochemical activities are distinct, both proteins function along the same general pathway.

Complementarity in the functions of Hsp104 and Hsp70 helps to explain why different species depend on different constellations of heat-shock proteins for thermotolerance¹⁷. *Drosophila*,

for example, does not even express an Hsp100 protein in response to heat; instead, it relies upon massive synthesis of Hsp70 directed by amplified genes (see ref. 17 for review). All organisms must: (1) balance the beneficial functions of particular heat-shock proteins against their detrimental effects^{8,18}; (2) use the heat-shock proteins that are most effective in protecting or repairing their own most critical cellular targets; and (3) develop a response appropriate for the types of stress they normally encounter. Thus, stress responses represent a balance between the biochemical advantages and disadvantages of the various heat-shock proteins within the context of a particular physiology and environment. □

Received 30 August; accepted 17 October 1994.

- Gottesman, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3513–3517 (1990).
- Parsell, D. A., Sanchez, Y., Stitzel, J. D. & Lindquist, S. *Nature* **353**, 270–273 (1991).
- Squires, C. L., Pedersen, S., Ross, B. M. & Squires, C. J. *Bact.* **173**, 4254–4262 (1991).
- Squires, C. & Squires, C. L. *J. Bact.* **174**, 1081–1085 (1992).
- Kruger, E., Volker, U. & Hecker, M. *J. Bact.* **176**, 3360–3367 (1994).
- Sanchez, Y., Taulien, J., Borkovich, K. A. & Lindquist, S. *EMBO J.* **11**, 2357–2364 (1992).
- Sanchez, Y. & Lindquist, S. *L. Science* **248**, 1112–1115 (1990).
- Sanchez, Y. et al. *J. Bact.* **175**, 6484–6491 (1993).
- Escher, A., O'Kane, D. J., Lee, J. & Szalay, A. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6528–6532 (1989).
- Schroder, H. T., Langer, T., Hartl, F.-U. & Bukau, B. *EMBO J.* **12**, 4137–4144 (1993).
- Webster, D. L. & Watson, K. *Yeast* **9**, 1165–1175 (1993).
- Skowyra, D., Georgopoulos, C. & Zyllicz, M. *Cell* **62**, 939–944 (1990).
- Shanen, S. L. et al. *Biochemistry* **21**, 5539–5551 (1982).
- Wickner, S., Hoskins, J. & McKenney, K. *Nature* **350**, 165–167 (1991).
- Wickner, S., Hoskins, J. & McKenney, K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7903–7907 (1991).

16. Langer, T. et al. *Nature* **356**, 683–689 (1992).
17. Parsell, D. A. & Lindquist, S. *Rev. Genet.* **27**, 437–496 (1993).
18. Feder, J. J., Rossi, J. M., Solomon, J., Solomon, N. & Lindquist, S. *Genes Dev.* **6**, 1402–1413 (1992).
19. Treco, D. A. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M., Brent, R. & Kingston, R. E.) 13.1.1–13.1.7 (Wiley, New York, 1989).
20. Kurtz, S., Gordon, E. & Lindquist, S. L. in *Sequence Specificity in Transcription and Translation* 611–620 (Liss, New York, 1985).
21. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
22. Wright, R. & Rine, J. in *Methods in Cell Biology* (ed. Tartakoff, A. M.) 473–512 (Academic, New York, 1989).
23. Byers, B. & Goetsch, L. *Meth. Enzym.* **194**, 602–608 (1991).
24. van Tuinen, E. & Riezman, H. J. *Histochem. Cytochem.* **35**, 327–333 (1987).
25. Schirmer, E., Lindquist, S. & Vierter, E. *Pl. Cell* (in the press).

ACKNOWLEDGEMENTS. We thank A. Szalay and A. Escher for the luciferase fusion gene and for luciferase-specific antiserum. Y. Sanchez, together with A.S.K. and S.L., made the original observation of aggregates in *hsp104* mutant cells. We thank E. Schirmer and J. Vogel for critical reading of the manuscript, and D. Wang for assistance in preparing figures.

The X-ray structure of a growth hormone–prolactin receptor complex

William Somers, Mark Utsch, Abraham M. De Vos & Anthony A. Kossiakoff*

Genentech Inc., Department of Protein Engineering,
460 Point San Bruno Boulevard, South San Francisco,
California 94080, USA

THE human pituitary hormones, growth hormone (hGH) and prolactin (hPRL), regulate a large variety of physiological processes, among which are growth and differentiation of muscle, bone and cartilage cells, and lactation¹. These activities are initiated by hormone–receptor binding. The hGH and hPRL receptors (hGH_R and hPRL_R, respectively) are single-pass transmembrane receptors from class 1 of the haematopoietic receptor superfamily^{2,3}. This classification is based on sequence similarity in their extracellular domains, notably a highly conserved pentapeptide, the so-called ‘WSXWS box’, the function of which is controversial. All ligands in class 1 activate their respective receptors by clustering mechanisms⁴. In the case of hGH, activation involves receptor homodimerization in a sequential process: the active ternary complex containing one ligand and two receptor molecules is formed by association of a receptor molecule to an intermediate 1:1 complex^{5–8}. hPRL does not bind to the hGH receptor, but hGH binds to both the hGH_R and hPRL_R, and mutagenesis studies have shown that the receptor-binding sites on hGH overlap⁹. We present here the crystal structure of the 1:1 complex of hGH bound to the extracellular domain of the hPRL_R. Comparisons with the hGH–hGH_R complex¹⁰ reveal how hGH can bind to the two distinctly different receptor binding surfaces.

As the extracellular domain of the hPRL_R only forms a stable 1:1 complex with hGH in solution, we used this complex in our crystallographic studies. The structure of this complex (Fig. 1a) was determined at 2.9 Å resolution and refined to a crystallographic *R*-factor of 0.22 (Table 1). The overall similarity with the hGH–hGH_R complex¹⁰ (Fig. 1b) can be seen in Fig. 1c, which shows the superposition of the structure of the hGH–hPRL_R complex on the equivalent 1:1 portion of the 1:2 hGH–hGH_R complex. The hGH molecule has the classical cytokine fold⁴, a four-helix bundle motif characterized by the first two helices running parallel to each other but antiparallel to the last two. The four-helix bundle structure of the hGH molecule in the hPRL_R complex superimposes well on that of the hGH_R (1:2) complex (0.82 Å r.m.s. on Cα positions). Larger

TABLE 1 Crystallographic data, refinement statistics and interfacial hydrogen bonds

Data collection statistics				
Dataset	Number of crystals	Number of observations	Number of reflections	R_{merge}
FAST	3	35,239	9,754	0.077
SSRL	6	17,794	8,293	0.095
Combined	9	53,033	9,883	0.108
Refinement statistics				
No. non-hydrogen atoms	3,123			
Disordered residues	hGH: 149–152, 191; hPRL _R : 1, 31–33, 84–86, 205–211			
Average <i>B</i> -factor	hGH: 55 Å ² ; hPRL _R : 42 Å ²			
Resolution range	10–2.9 Å			
<i>R</i> -value (all data)	0.22			
r.m.s. bond length deviation	0.019 Å			
r.m.s. bond angle distance deviation	0.046 Å			
r.m.s. difference in <i>B</i> -factor	0.63 Å ² (main chain atoms), 0.66 Å ² (side-chain atoms)			
Intermolecular hydrogen bonds in hGH–hPRL _R and hGH–hGH _R				
hGH	hPRL _R	hGH	hGH _R	
Ser 62 O	Met 103 N	Pro 61 O	Ile 103 N	
Lys 168 N ζ	Trp 104 O	Lys 168 N ζ	Trp 104 O	
Ser 51 O γ	Glu 75 O ϵ 2	Lys 41 N ζ	Glu 127 O ϵ 2	
Tyr 164 O η	Glu 75 O ϵ 1	Arg 167 N η 1	Glu 127 O ϵ 1	
Arg 167 N η 2	Asp 124 O δ 2	Arg 167 N η 2	Glu 127 O ϵ 1	
Arg 178 N η 1	Thr 171 O γ 1	Arg 178 N η 2	Ile 165 O	
Arg 178 N η 1	Phe 170 O	Thr 175 O γ 1	Arg 43 N η 1	
Arg 178 N η 2	Gln 193 O ϵ 2	Gln 46 N ϵ 2	Glu 120 O ϵ 2	
Asp 171 O δ 1	Tyr 127 O η	Asp 171 O δ 2	Arg 43 N η 2	

hGH and hPRL_R were purified, complexed and crystallized as described¹⁶. Crystals were in space group *P*₂₁₂₁₂, with *a* = 154.2 Å, *b* = 69.8 Å, *c* = 43.4 Å. Data were collected at room temperature from three crystals using an Enraf-Nonius FAST area detector mounted on a Rigaku RU200 rotating anode generator and monochromated CuKα X-rays, and processed using MADNES¹⁷ and PROCOR¹⁸. The program X-PLOR¹⁹ was used for molecular replacement with the 1:1 hGH–hGH_R structure (A.M.d.V., unpublished results) as a search model. The rotation function gave no clear solutions, but Patterson correlation (PC) refinement²⁰ of the rotation function peaks treating the hormone and both receptor domains as independent rigid bodies gave a solution with a PC value of 0.13, clearly above the highest noise peak at 0.07 (data between 15 and 4 Å resolution). The translation search gave a correlation value of 0.36 with the highest noise peak at 0.24 and a crystallographic *R*-value of 0.49, decreased to 0.47 by rigid-body refinement (15–3.5 Å). The refinement dataset was generated by merging the molecular replacement dataset with data collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on a MAR image plate scanner and processed with XDS¹⁸, giving an overall completeness of 93% (89% between 3 and 2.9 Å). The structure was refined using a combination of simulated annealing and positional refinement with X-PLOR and manual intervention using FRODO²¹. The final stages of refinement were performed using PROLSQ²² with highly-restrained individual temperature factors, resulting in an *R*-value of 0.22 (10–2.9 Å).

differences are found in the loop structures and in two small mini-helices. The loop connecting helices 2 and 3 (residues 92–111) does not contact the receptor and generally is not well ordered. The first turn in helix 3 is partially unwound compared to the structure of the hGH–hGH_R complex; a similar unwinding occurs in the structure of a free hGH variant¹¹. The loop connecting helices 3 and 4 is also poorly ordered, but was modelled into weak electron density in a similar conformation as observed in the 1:1 hGH_R complex. The complexes also differ at mini-helix 1 (M1 in Fig. 1, residues 42–46), in which the N-terminal turn is partially unwound, and in the segment of loop immediately following it. Because changes in this region have also been seen in the structure of the hGH variant, where it

* To whom correspondence should be addressed.

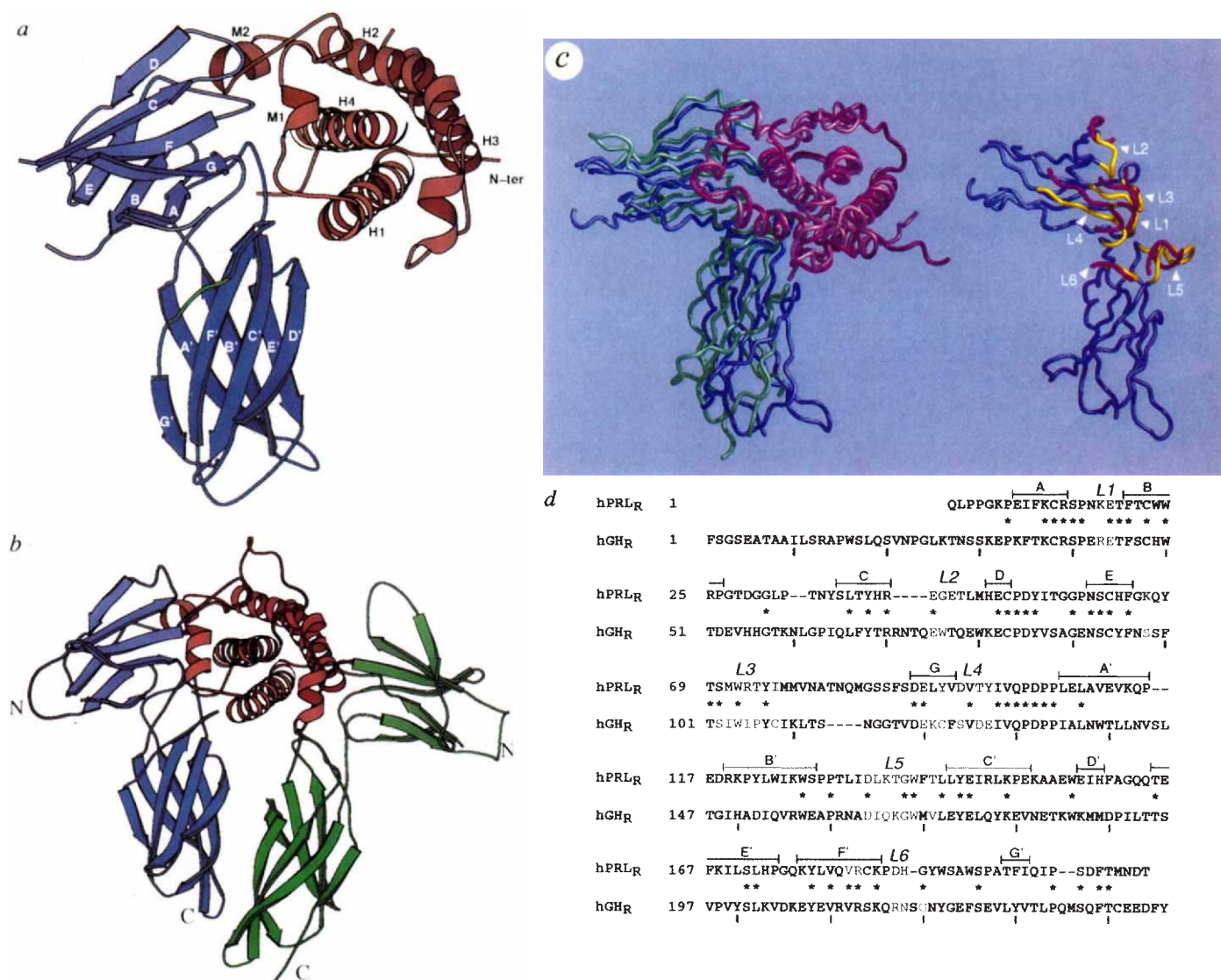


FIG. 1 *a*, Ribbon representation of the hGH-hPRL_R complex, with the hormone in red and the receptor in blue. The four helices making up the core of hGH are labelled H1 through H4; small additional helical segments are denoted as M1 and M2. The strands in the two receptor domains are labelled alphabetically, A-G for the N-terminal domain and A'-G' for the C-terminal domain. Residues making up the WSXWS-box are in green. Segments of the polypeptide chain that could not be seen in the electron density maps have been omitted. *b*, Ribbon rendering of the 1:2 hGH-hGHR complex¹⁰. This and *a* were made with the program MOLSCRIPT²³. *c*, Left, overlay of the hGH-hPRL_R complex and the hGH-hGHR complex, after superposition of the four-helix bundle core of hGH.

was involved in crystal packing interactions, this segment can apparently readily adapt to its environment¹¹.

Even though the sequences of the hPRL_R and hGHR share only 28% identity, their structures are very similar (the structure-based sequence alignment is shown in Fig. 1*d*; the residue numbering of hGHR is used). The hPRL_R and hGHR extracellular portions consist of a tandem repeat of fibronectin type III-like modules connected by a linker of five residues¹⁰. Each module is very similar between the hPRL_R and hGHR: both modules can be superimposed with an r.m.s. difference of about 0.85 Å between equivalent β-strand Cα positions. The N-terminal domain contains a pair of disulphide bonds that are highly conserved in the haematopoietic superfamily; strands A and B are bridged by Cys_R38 and Cys_R48, and D and E by Cys_R83-Cys_R94. The WSXWS box is found immediately preceding strand G' of the C-terminal domain (Fig. 1*a*).

hGH from the complex with hPRL_R (blue) is red, and hGH from the complex with the hGHR (green) is magenta. Right, superposition of the six binding loops of the hPRL_R (yellow) and of the hGHR (red) on the structure of the hPRL_R (blue). This and similar figures were made with the graphics program MIDAS²⁴. *d*, Sequence alignment of the hPRL_R and the hGHR. Contact residues are blue (hPRL_R) and green (hGHR). Strands are labelled A to G for the N-terminal domain and A' to G' for the C-terminal domain. Binding loops are denoted L1 to L6. Sequence identities are designated by an asterisk. In the text, hGHR numbering is used.

An important difference between the hPRL_R and hGHR complexes is that both the N-terminal and the C-terminal receptor domains are translated and rotated with respect to each other (Fig. 1*c*). The differences are 1.9 Å and 8°, and 5.2 Å and 8°, respectively, for the N-terminal and C-terminal domains. Nevertheless, the hormone-receptor interface in both 1:1 complexes consists of essentially identical structural epitopes of hGH in contact with virtually the same receptor interface surfaces. This is shown in Fig. 2, which compares the buried surface area in both 1:1 hGH-receptor complexes, in which roughly the same total surface areas are buried on each receptor (~1,350 Å²). The principal contact residues on the hormone are contained in the loop connecting helices 1 and 2, and in helix 4. The ligand-receptor interface observed in the crystal structure contains the 'functional epitope' deduced from mutagenesis studies⁹ but is significantly larger. Most of the functionally important residues

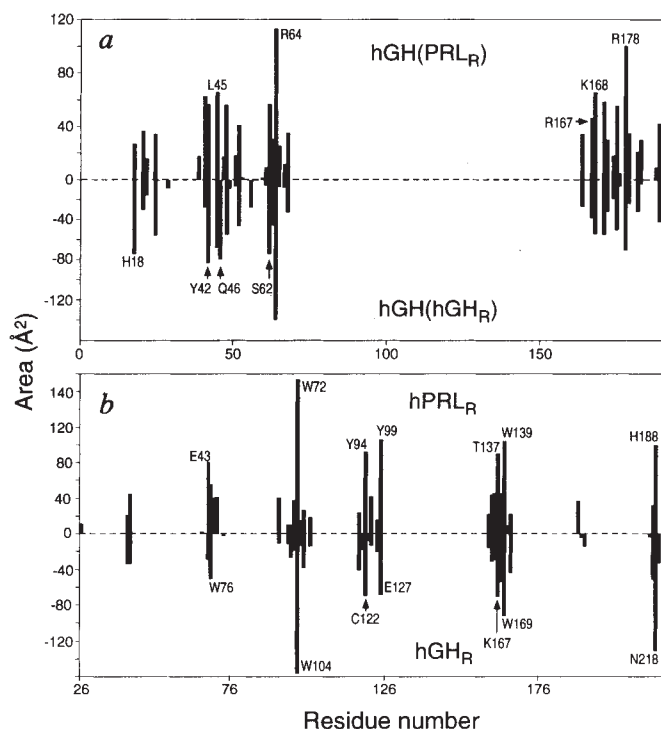


FIG. 2 Surface area buried in the hormone-receptor interface. *a*, Residues on hGH in the complex with the hPRL_R (top) and with the hGH_R (bottom), respectively. *b*, Residues on the hPRL_R (top) and the hGH_R (bottom). Both hPRL_R and hGH_R numbering is used, but in the text all sequences are referred to using the hGH_R numbering.

of hGH contribute to the binding affinity for both hPRL_R and hGH_R, but there are some, denoted specificity determinants, that are critical for binding to one of the receptors and not the other^{9,12}, most importantly Arg 167 and Glu 174.

The binding determinants of the receptor are contained on six loops that extend out from the β -sheet core and contain between two (loops L1 and L6) and eight (loops L4 and L5) interfacial residues (Fig. 1*d*). The superposition of the hGH-hPRL_R complex onto the hGH-hGH_R shows that there are large differences in the position of most of these loops; for instance, residues in the main chain of L4 differ on average by 3.5 Å and L5 residues vary by about 2.5 Å. However, comparison of the individual loops themselves indicates that they have similar main chain conformations in hPRL_R and hGH_R; for instance, superposition of the main chain of loop L4 in the two receptors alone gives an r.m.s. deviation of about 0.9 Å. Thus, the differences in the interface are caused by rigid-body changes in domain orientation; these changes are triggered by differences in the interactions between the domain-domain linker (residues 127-131) and the hormone. It is notable that these interactions include Arg 167, one of the specificity determinants of hGH⁹. Position 127 of the receptors is followed by seven residues that are identical in the hPRL_R and hGH_R. However, residue 127 itself is tyrosine in the hPRL_R and glutamate in the hGH_R. A compensatory shift of the linker places the side chain of residue 127 in a different environment with respect to the hormone in each complex (Fig. 3*a*). In the hGH-hGH_R complex, Glu_R 127 hydrogen bonds to Arg 167 of the hormone, whereas in the hPRL_R complex the Tyr_R 127 hydroxyl interacts with hGH Asp 171 while the ring packs against the hGH Arg 167 methylene groups. Arg 167 is still involved in a salt bridge, but now to the carboxylate of Asp_R 124, which in turn hydrogen bonds to Thr_R 106. Mutagenesis has shown that the R167A mutation in hGH does not affect binding affinity for the hGH_R, but decreases binding to the hPRL_R by 770-fold⁹. Apparently, in the hGH-hGH_R complex, the enthalpy gain of forming the salt bridge is offset by an

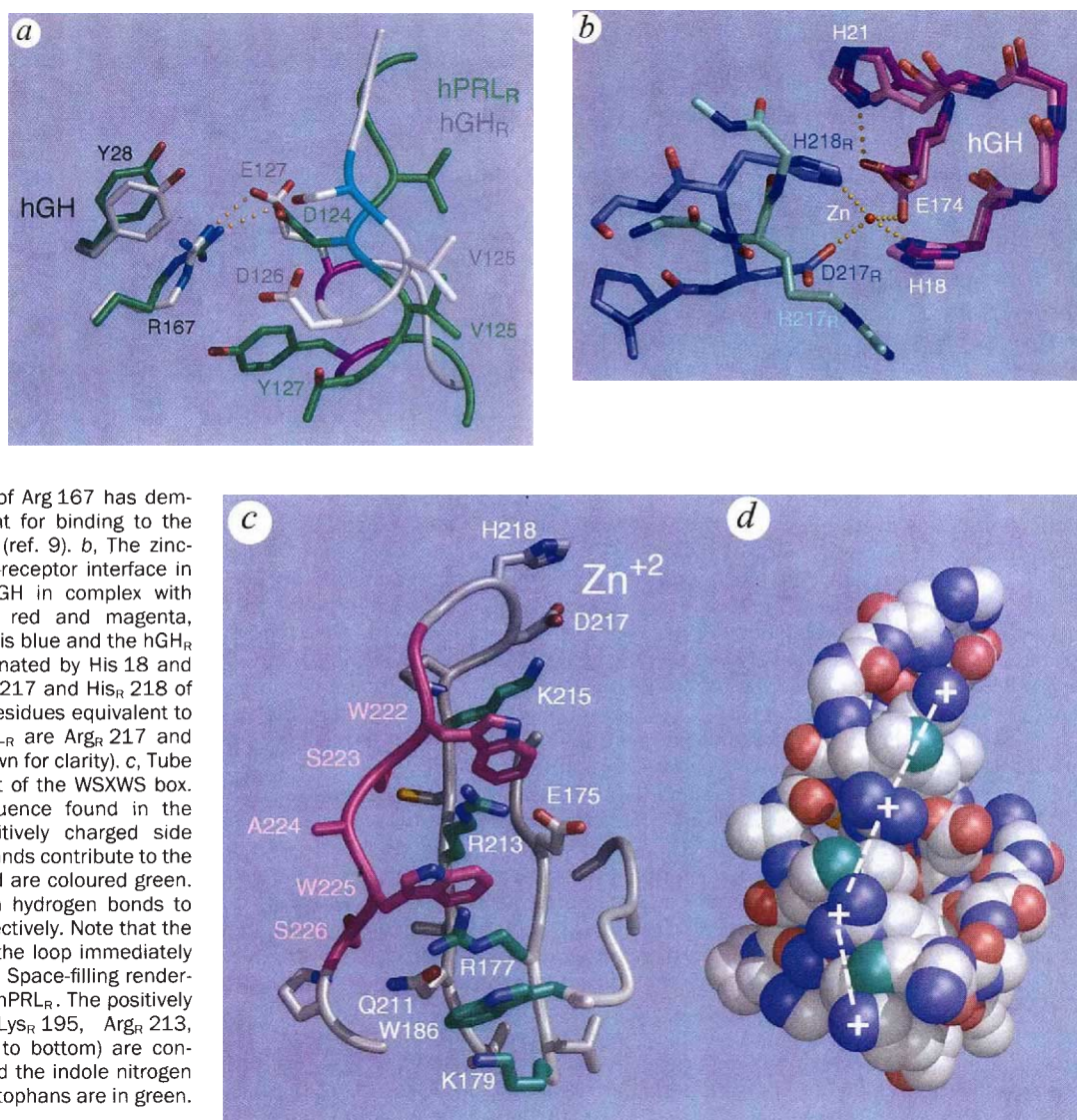
entropy loss incurred by a reduction in side-chain flexibility compared to the uncomplexed state; in addition, the salt bridge is readily accessible to bulk solvent. In the free hPRL_R, the side chain of Asp_R 124 is probably already pre-oriented by its interaction with Thr_R 106, whereas the salt bridge is screened from solvent by the side chain of Tyr_R 127.

The global shifts of the two domains require a number of adjustments to maintain important interactions, some subtle, others more significant. Two tryptophans in hGH_R, Trp_R 104 and Trp_R 169, are essential for hormone binding in the hGH-hGH_R complex (refs 9, 13; T. Clackson and J. Wells, personal communication). In hPRL_R these tryptophans are conserved and bind in nearly identical hydrophobic environments as those in the hGH_R complex; these environments are formed by hormone residues that were identified as important binding determinants by alanine scanning¹⁴. In contrast to the similar binding of the tryptophans are the differences observed for the intermolecular hydrogen bonds and salt bridges (Table 1). Of nine hormone-receptor salt bridges or hydrogen bonds, only Lys 168 N ζ to Trp_R 104 O is conserved, whereas a second hydrogen bond between main-chain atoms is very similar (Ser 93 O to Met_R 103 N). Sequence or positional differences between the receptors result in formation of alternative hydrogen bonds. An example is Asp 171: in the hGH-hGH_R complex, the carboxylate interacts with Arg_R 43, but in the hGH-hPRL_R complex, Lys_R 43 is too short to form a favourable interaction with Asp 171 (an observation consistent with mutagenesis data showing that the K43A mutation has little effect on binding¹⁵). Instead, Asp 171 reorients and interacts with the hydroxyl of Tyr_R 127.

A unique feature of the binding interaction between hGH and hPRL_R is at loop L6, where the sequences differ: hGH_R has (starting at position 217) Arg, Asn, Ser, whereas hPRL_R has Asp, His, deletion. In the hGH-hPRL_R complex, the interaction is accommodated by the presence of a zinc-binding site that potentiates a 10,000-fold Zn²⁺ dependence toward receptor binding¹². This site is formed by two ligands on the receptor (Asp_R 217 and His_R 218) and two on the hormone (His 18 and Glu 174). Glu 174 also hydrogen bonds to the N δ 1 of His 21 through its other carboxylate oxygen (Fig. 3*b*). The observed interactions allow us to reinterpret the mutagenesis data, which showed that the histidines and the glutamate on the hormone and the receptor histidine are important for maintaining tight binding for the hPRL_R but not the hGH_R^{9,12}. Histidine 21 is not a direct ligand to the zinc but probably serves to pre-orient the side chain of Glu 174. Furthermore, we predict that mutation of Asp_R 217 to alanine would affect zinc binding and hence diminish the affinity of the resulting mutant for hGH.

The structure of the hPRL_R shows that the Trp-Ser-Ala-Trp-Ser sequence corresponding to the WSXWS box forms a stretch of irregular extended chain immediately preceding strand G', quite removed from the ligand-binding interface (Fig. 1*a*). The alanine side chain points out into the solvent, whereas the two tryptophan side chains fold back on top of the solvent-accessible side of the four-stranded sheet (Fig. 3*c*). The two serines appear to have a structural role because their side chains form hydrogen bonds to main-chain atoms of the neighbouring strand: Ser_R 223 O γ to Cys_R 214 N and Ser_R 226 O γ to Val_R 212 N. The WSAWS box is part of a much larger, highly organized pattern covering much of the surface of the four-stranded sheet and consisting of alternating hydrophobic and charged groups (Fig. 3*c*). The pattern starts from the bottom with Lys_R 179 followed by Trp_R 186, Arg_R 177, Trp_R 225, Arg_R 213, Trp_R 222, Lys_R 215. Additional contributions to this motif come from two hydrogen bonds, one from the carboxylate of Glu_R 175 to Arg_R 213 N η 1. The disposition of the participating side chains creates an intriguing electrostatic surface. Each of the three tryptophans is oriented so that the hydrophobic portion of the side chain is fully buried, but leaving the indole ring nitrogen atom accessible to solvent (Fig. 3*d*). Together with the charged termini of the

FIG. 3 a, Hormone-receptor interactions involving the domain-domain linker in the receptor, with the hGH-hPRL_R complex in green and the hGH-hGH_R complex in white. The receptor linker region (at the right-hand side) is translated by about 3 Å, resulting in a displacement of Asp 124 of the hPRL_R compared to the equivalent residue, Ser 124, of the hGH_R. The result is that Arg 167 of the hormone interacts with Asp 124 of the hPRL_R, but with Glu 127 rather than Ser 124 of the hGH_R. Alanine mutagenesis of Arg 167 has demonstrated that it is important for binding to the hPRL_R, but not to the hGH_R (ref. 9). b, The zinc-binding site in the hormone-receptor interface in the hGH-hPRL_R complex. hGH in complex with hPRL_R and with hGH_R are red and magenta, respectively, while the hPRL_R is blue and the hGH_R green. The zinc ion is coordinated by His 18 and Glu 174 of hGH and by Asp_R 217 and His_R 218 of the hPRL_R. In the hGH_R the residues equivalent to the zinc ligands of the hPRL_R are Arg_R 217 and Asn_R 218 (side chain not shown for clarity). c, Tube rendering of the environment of the WSXWS box. The Trp-Ser-Ala-Trp-Ser sequence found in the hPRL_R is in magenta. Positively charged side chains from neighbouring strands contribute to the extended surface pattern and are coloured green. Glu_R 175 and Gln_R 191 form hydrogen bonds to Arg_R 213 and Arg_R 177, respectively. Note that the zinc ligands are situated on the loop immediately preceding the WSXWS-box. d, Space-filling rendering of the WSXWS box in the hPRL_R. The positively charged end groups of Lys_R 195, Arg_R 213, Arg_R 177 and Arg_R 179 (top to bottom) are connected with a dotted line, and the indole nitrogen atoms of the interleaving tryptophans are in green.



two arginines and the two lysines, a large corridor of positive charge is produced in which the interleaving tryptophans separate the successive charges by about 8 Å. Sequence alignment of the receptors in the superfamily shows that analogous patterns of hydrophobic side chains interleaved by hydrogen-bonding groups are probably present in all class I receptors and suggests that the folding, as well as the spatial arrangement, of the β -strands of the C-terminal domain are very similar for all these receptors.

The comparison between hGH_R and hPRL_R complexes reveals an extraordinary adaptability of the receptors' fibronectin type-III domains to form competent binding epitopes.

There appear to be no specific rules governing molecular recognition processes, only general ones. Binding and specificity determinants are generally separable; binding is influenced by the steric and hydrophobic character of the interface, specificity by hydrophilic interactions that affect the orientation of the receptor domains. Modelling studies suggest that the differences in domain orientation are biologically important: the 1:1 hPRL_R complex seen in the crystal represents the 'productive' template, preset and ready to bind a second hPRL_R molecule to form the active complex, but sterically unable to form a 'mixed' ternary complex by accommodating hGH_R as a second receptor molecule. □

Received 17 August; accepted 27 September 1994.

- Chawla, R. K., Parks, J. S. & Rudman, D. A. *Rev. Physiol.* **34**, 519–547 (1983).
- Bazan, J. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6934–6938 (1990).
- Cosman, D. *Cytokine* **5**, 95–106 (1993).
- Sprang, S. R. & Bazan, J. F. *Curr. Opin. Struct. Biol.* **3**, 815–827 (1993).
- Ullsch, M., De Vos, A. M. & Kossiakoff, A. A. *J. molec. Biol.* **222**, 865–868 (1991).
- Cunningham, B. C. *et al. Science* **254**, 821–825 (1991).
- Fuh, G. *et al. Science* **256**, 1677–1680 (1992).
- Fuh, G., Colosi, P., Wood, W. I. & Wells, J. A. *J. biol. Chem.* **268**, 5376–5381 (1993).
- Cunningham, B. C. & Wells, J. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3407–3411 (1991).
- De Vos, A. M., Ullsch, M. & Kossiakoff, A. A. *Science* **255**, 306–312 (1992).
- Ullsch, M., Somers, W., Kossiakoff, A. A. & De Vos, A. M. *J. molec. Biol.* **236**, 286–299 (1994).
- Cunningham, B. C., Bass, S., Fuh, G. & Wells, J. A. *Science* **250**, 1709–1712 (1990).
- Bass, S. H., Mulkerrin, M. G. & Wells, J. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4498–4502 (1991).

- Cunningham, B. C. & Wells, J. A. *J. molec. Biol.* **234**, 554–563 (1993).
- Rozakis-Adcock, M. & Kelly, P. A. *J. biol. Chem.* **267**, 7428–7433 (1992).
- Ullsch, M. & De Vos, A. M. *J. molec. Biol.* **231**, 1133–1136 (1993).
- Messerschmidt, A. & Pflugrath, J. W. *J. appl. Crystallogr.* **20**, 306–315 (1987).
- Kabsch, W. *J. appl. Crystallogr.* **21**, 916–934 (1988).
- Brünger, A. T. *X-PLOR Reference Manual* (Yale Univ. Press, New Haven, 1992).
- Brünger, A. T. *Acta crystallogr.* **A46**, 46–57 (1990).
- Jones, A. T. *J. appl. Crystallogr.* **11**, 268–272 (1978).
- Hendrickson, W. A. *Meth. Enzym.* **115**, 252–270 (1985).
- Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1991).
- Ferrin, T. E., Huang, C. L., Jarvis, L. E. & Langridge, R. *J. molec. Graph.* **6**, 13–27 (1988).

ACKNOWLEDGEMENTS. We thank B. Snedecor and H. Chen in the fermentation department at Genentech, M. Soltis and H. Luecke at the Stanford Synchrotron Radiation Laboratory for help with data collection, T. Clackson, H. Lowman and J. Wells for discussions and for making available unpublished data and K. Andow for graphics assistance.

Forewarned is four-armed

A new solution structure of the oligomerization domain of p53 provides a framework for interpreting a wealth of biochemical data but casts doubt on details of an earlier structure.

THE nuclear phosphoprotein p53 has progressed from its first identification as a lowly tumour-associated antigen through oncoprotein to the position it holds today as one of the central suppressors of cancer development. Around half of all human cancers involve p53 dysfunction, resulting most frequently from mutation and/or deletion of the *p53* gene itself, or from the binding and inactivation of p53 by an oncoprotein. We have a wealth of information about p53's activities, but details of its three-dimensional structure remained elusive until the publication of the structures of two of its domains last July in *Science*^{1,2} and of another view in this month's *Nature Structural Biology*³.

p53 is a protein of 393 amino acids which acts as a tumour suppressor by controlling the progression of cells through the cell cycle. If a cell's genome becomes damaged, the cycle is halted at the G1 stage by the interception of p53, until either the damage can be repaired or, failing that, the cell is programmed for destruction. This activity, which has earned p53 the title 'guardian of the genome', is accomplished not by arresting the cell cycle directly but by stimulating transcription of growth-inhibitory genes and possibly suppressing the transcription of other genes by an interaction with the TATA-box-binding protein.

The primary structure of p53 can be divided into three domains, each responsible for a different function. The amino terminus is concerned with transcriptional activation, the central core (residues 102–292) forms a DNA-binding domain, and the carboxy-terminal third of the protein, which contains several regulatory phosphorylation sites, is instrumental in the formation of p53 tetramers, the active form *in vivo*.

Truncated p53 containing only the oligomerization domain induces p53 dysfunction in the presence of intact p53 (ref. 4). This 'minimum transforming domain' (residues 302–360) forms hetero-oligomers with the wild-type protein. Presumably these p53 hetero-oligomers are non-functional, explaining why some

p53 mutations produce a dominant negative phenotype.

The structure of a peptide fragment that overlaps the minimal transforming domain, p53tet (residues 303–366), has now been determined using NMR spectroscopy³. This structure reveals a tetramer with the same general fold for the individual monomers as that already found for a shorter peptide (residues 319–360)², but the quaternary arrangement of the chains is very different.

Only the core of each monomer is well defined: 21 residues at the N terminus and ten residues at the C terminus appear to be in a flexible, random-coil configuration. The central section forms a relatively compact globular fold consisting of a β -strand, a turn and an α -helix. Two peptides come together as a dimer with their β -strands forming a two-stranded β -sheet. The pair of dimers associate with their α -helices apposed and their β -sheets at opposite sides of what is essentially a four-helix bundle. It is in the packing of this dimer of dimers that the two structures^{2,3} differ: in p53tet the apposed helices from opposite dimers are in a parallel orientation, whereas in the earlier structure the dimers are rotated by 37°, resulting in an antiparallel arrangement. Associated with this rotation is a difference in the α -helices, which bend back towards the β -sheets in p53tet and outward towards the opposite dimer in the other structure.

Derivation of the structure of an oligomer of identical subunits by NMR spectroscopy is particularly hard, relying as it does on the determination of distance constraints between individual atoms based on the assignment of peaks in the NMR spectrum (NOEs) and the need to distinguish inter- and intra-subunit NOEs between identical residues. Indeed, the variations between the two oligomerization-domain structures stems from different assignments of inter-subunit NOEs. These assignments are usually made by mixing labelled and unlabelled proteins, removing inter-subunit NOEs and simplifying the spectrum, an approach used by Clore and colleagues². For determination of the p53tet structure, Lee *et al.* subtly modified this procedure³.

An equimolar mixture of p53tet was made under native conditions so that initially all tetramers were either entirely ¹³C-labelled or unlabelled. Dimers exchanged quickly, so differences between the spectra of mixed and unmixed protein

identified inter-dimer NOEs. With time the spectrum changed, owing to subunit exchange within the dimer. By this method not only were intra-subunit, intra-dimer and inter-dimer NOEs identified, but the p53 dimer was shown to be more stable than the complete tetramer.

How do these structures relate to the entire p53 molecule and its biological activity? The consensus DNA-binding site for p53 consists of two palindromic 10-base-pair sequences which can be separated by up to 13 base pairs⁵. Each palindrome could interact with two DNA-binding domains, as indicated by the structure of the DNA-binding domain in complex with DNA¹. In p53tet the N-terminal portions of the peptides are so arranged that this flexible DNA binding can be achieved without distorting the DNA³. In contrast, the structure of Clore *et al.* suggests that DNA binding will induce a considerable bend².

The structures also have implications for the control of p53 function. The hair-pin shape of the individual subunits allows close positioning of the DNA-binding domain and the C terminus, where phosphorylation of serine 393 stimulates the DNA-binding activity of p53. Deletion of this region of the protein or binding of a specific antibody to it also promotes DNA binding, so an interaction of the C terminus with the DNA-binding domain may obstruct binding of DNA.

Many tantalizing questions remain, not least the discrepancy between the two oligomerization domain structures. It is formally possible that this protein fragment can exist in two conformations dependent on the conditions, although there are no examples of the same amino-acid sequence occurring in such different conformations. It is more likely that only one is correct, and the more flexible DNA-binding model and an as-yet unpublished crystal structure favour that of Lee and colleagues over the earlier structure.

Christopher Surridge

Christopher Surridge is Assistant Editor of Nature Structural Biology.

Also in this month's *Nature Structural Biology*: designing membrane proteins; DNA binding of the *ets* domain; structural conservation in early immunoglobulin evolution; inverted peptide binding to a GRB2 SH3 domain; racemic trichogin A IV; and galectin networks in crystals.

1. Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P. *Science* **265**, 346–355 (1994).
2. Clore, G. M. *et al.* *Science* **265**, 386–391 (1994).
3. Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D. & Arrowsmith, C. H. *Nature struct. Biol.* **1**, 877–890 (1994).
4. Shaulian, E., Zauberman, A., Ginsberg, D. & Oren, M. *Molec. cell. Biol.* **12**, 5581–5592 (1992).
5. El-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. *Nature Genet.* **1**, 45–49 (1992).

Lasers — in the spotlight

The latest in laser technology this week features a new polarized, frequency stabilized HeNe laser system, matrix-assisted, laser-desorption ionization, time-of-flight mass spectrometers and a series of pyroelectric joulemeters.

FROM Fritsch — the 'analysette 22' Economy laser particle sizer (*Reader Service No. 100*). The instrument can be used to determine particle size distributions of solids in suspension and emulsions in the range of 0.1–600 μm . Design features include an



Economy laser particle sizer for measuring in the 0.1–600 μm range.

accelerated measuring cycle lasting about a minute, up to 310 measuring channels, an automatic device for 'correct' measuring range adjustment for samples of unknown distribution, Microsoft Windows user interface, printout of the results using the 'report generator', data exporting using dynamic data exchange, output of statistical values, etcetera.

■ Types of laser

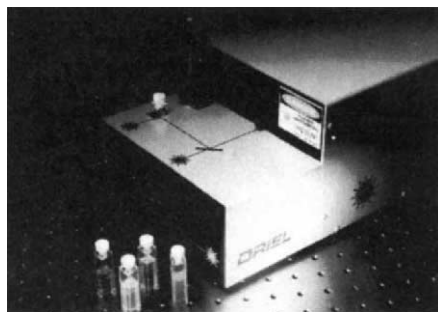
Melles Griot has a new polarized, frequency stabilized HeNe laser system (*Reader Service No. 101*). The laser has a stated long-term frequency stability of ± 2 MHz over 8 h of continuous use, and an intensity stability of up to ± 0.1 per cent. The system has a 'turnkey' locking feature for stabilization of frequency and intensity. Stated uses include inspection in applications such as interferometry, metrology, microscopy, spectroscopy and velocimetry.

The performance of Coherent's OPA laser system has been extended to shorter wavelengths and pulse durations (*Reader Service No. 102*). The new capabilities demonstrated recently from a 250 kHz mode-locked titanium:sapphire Mira/regenerative amplifier RegA 9000/OPA system are designed to enhance applications in the biosciences, chemistry and physics. These

NATURE • VOL 372 • 1 DECEMBER 1994

advances, Coherent says, have been made possible because of the energy, stability and beam quality of the CW ion-pumped system. These characteristics have allowed the observation of efficient frequency-doubling of the OPA 9400 signal pulses — from 480–700 nm in the visible range to 240–350 nm in the ultraviolet — with pulse energies greater than 10 nJ. In addition, the visible OPA signal pulses have been dispersion-compensated with a prism sequence to pulsewidths as short as 70 femtoseconds, producing, says Coherent, sufficient peak power to generate a white-light continuum. Further dispersion compensation of this continuum spectrum is said to produce widely tunable, visible pulses with a duration of less than 30 femtoseconds.

LOT-Oriel offers a new line of pulsed ultraviolet nitrogen lasers which utilize a 'hard seal' technology developed by the company's laser research group (*Reader Service No. 103*). This hard seal is designed to provide a longer tube life — of up to 10(8) shots. The line up includes a high-pressure 600 ps pulse laser and a higher energy 5 ns pulse version. Also available — a tunable



LOT-Oriel's pulsed nitrogen and dye lasers.

dye laser module that attaches to the output of either nitrogen laser; it is tunable from 350 to 750 nm. The dye module has two output ports, so it is possible to switch between nitrogen and dye output wavelengths. Oriel offers both lasers separately or as a complete package.

■ MALDI systems

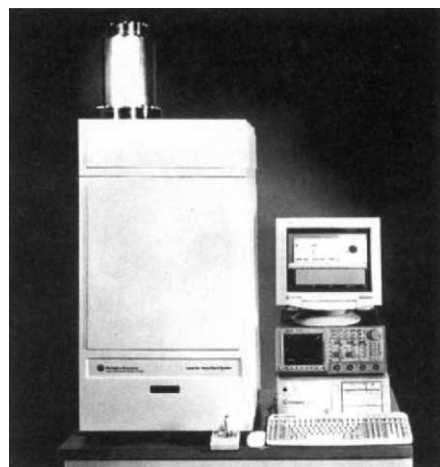
Shimadzu's Compact MALDI I molecular weight analyser combines matrix-assisted, laser-desorption ionization (MALDI) with a time-of-flight mass measurement system (*Reader Service No. 104*). Users can select from a wide range of sample substrates to provide optimum results from a variety of differing sample types, ranging from synthetic industrial polymers to protein enzyme digests. Flexibility is provided by the con-



Compact MALDI I system for molecular weight analysis.

tinuous scan mode when handling PVDF blots, HPLC and CZE fractions. The system, which takes up 2.6 square feet of bench space, is supplied with a computer, a range of applications software, a sample preparation unit and an installation kit. The Windows-based software includes control, processing and diagnostic programs, as well as packages for peptide/protein analysis and polymer calculations.

The Vestec LaserTec BenchTop II fully-automated reflector, MALDI time-of-flight mass spectrometer, available from GSG Interion, features a new vertical reflector for studying fragmentation and protein mixtures (*Reader Service No. 105*). In addition to high-resolution molecular weight information on peptide mixtures under 12,000 daltons, applications have been extended to the study of metastable decompositions. Spectra obtained with this configuration may provide sequence information on proteins and peptides directly, without the use of chemical modification or digestion. Designed



LaserTec BenchTop II for studying fragmentation and protein mixtures.

ADVERTISEMENTS

DNA SEQUENCING

Using an ABI 373
Automated Sequencing
System

£45 per run

- PCR products
- M13 inserts
- plasmid inserts
- single stranded DNA
- double stranded DNA
- 300-400 bases of sequence

Please contact us for more information

Molecular Medicine Unit
King's College School of Medicine & Dentistry
Tel: 071- 346 3126 Fax: 071-733 3877

Reader Service No.11

Biotin and Digoxigenin Label
Now controlled Label
guarantees enhanced sensitivity

TIB
MOLBIOL

Primers and Probes
Custom Synthesis for
Research and Diagnosis

Order by Fax
Germany +49 30-78 79 94 99
Italy +39 10-362 19 38

Reader Service No.25

O₂/CO₂ RESPIROMETER

Measures O₂ Consumption and CO₂ Production (BOD Rate) in Headspace of 1-80 Chambers. Sensitivity 0.2 µl/h. Fully Computerized. Optional Methane Sensor. Chambers headspace volume: 50 ml to 10 l.

Applications:
Bioremediation of soil or water, bacterial or fungal growth, oxidation of food, respiration of insects, animals or plants.

For More Information Contact:
COLUMBUS INSTRUMENTS
P.O. Box 44049
Columbus, OH 43204-0049 USA
Ph: (614) 276-0861 Fax: (614) 276-0529

Reader Service No. 4

PEPTIDE and DNA

Peptide and DNA engineering specialist:
Phosphoryl, Thiophosphoryl, Biotinyl,
Chloromethyl, Cyclic, Dimeric, Partial protected,
FITC, DNP, pNA, Phosphothionate and more ...

IonSpray MS and LC/MS

Synthetic or natural compounds analysis:
Organic, inorganic, or coordination compounds.
Drug metabolites, Enzyme digests, etc.
\$30/sample (MS), \$250/sample (LC/MS)

Call for more information (800-597-7873) or
Fax us sequence for quotation (510-371-1156).
PeptidoGenic Research & Co. Livermore, CA

Reader Service No.3

for rapid sample throughput and routine use, the LaserTec product line uses a video monitor for viewing each sample and a joystick for fast adjustment of sample position on a 100-sample plate on an x,y stage. The instrument is suited for the automatic analysis of a range of biomolecules, and is supplied in a benchtop cabinet with a 1.2 m vertical linear flight tube and a 2.6 m reflector.

In addition to determining molecular weights, Finnigan MAT's Vision 2000 MALDI time-of-flight mass spectrometer can be used to gain structural information as well, especially about sensitive biomolecules (Reader Service No. 106). Measurement of post source decay products (PSD) in MALDI is said to reveal useful sequence information of biopolymers such as peptides, glycan and nucleotides. Site determination of post-translational modifications in proteins, as well as modified bases in nucleotides, are considered to be prime applications of PSD.

■ Peripherals

Optilas is making available a range of **pyroelectric laser/energy meters** made by Ophir Optonics (Reader Service No. 107). The PE series of pyroelectric meters can be used to measure laser energies independently of either pulse duration or pulse repetition frequency. Compatible with Ophir's latest Nova display unit, PE detectors now allow laser users to measure both short and long pulses (up to 5 ms) at pulse repetition rates up to 5 kHz. The stated sensitivity of the detectors is 1 µJ. When used with the Nova display, the system can monitor and record pulse energies, with the results displayed in histogram format.

Hitachi Denshi's new **He-Ne laser CCD camera** has been designed to overcome some of the problems with conventional CCD cameras—in particular, interference fringes of 15–30 per cent generated by the He-Ne laser reflected in the camera (Reader Service No. 108). With internal fringe suppression,



Hitachi Denshi's He-Ne CCD laser camera.

the interference fringes in this camera are said to be reduced to around 1.5 per cent. The stated horizontal resolution is 570 TV lines and the spectral response is from 400 to 1,000 nm (25 per cent video output), with the maximum interference suppression at 632.8 nm.

AG Electro-Optics is distributing ILX Lightwave's improved LWM6500B **laser wavelength meter** (Reader Service No. 109). This instrument is designed to measure wavelengths from 500–1,600 nm with a stated accuracy and resolution of ± 0.5 nm and 0.01 nm, respectively. The lower wavelength range of 500–1,000 nm uses a silicon sensor head; InGaAs is used for 1,040 to 1,600 nm. Applications include dye-laser monitoring, laser diode testing, laser diode mode hopping studies and frequency mixing experiments. An IEEE 488 interface is optional and a free LabView driver is available.

The new Super BeamAlyzer **laser beam profiler** that can be obtained from Melles Griot provides a single test station for analysing the parameters of laser beams up to 9 mm in diameter (Reader Service No. 110).



Laser beam profiler — see Melles Griot.

The device can be used simultaneously to measure the width, profile, position, power and power noise spectral density of a continuous wave laser beam. One of the Super BeamAlyzer's special features is its ability to analyse the power noise and modulation spectrum of a laser in the 100 Hz-to-6 MHz frequency range. Enhanced spatial resolution is provided with seven scanning knife-edges using tomographic reconstruction techniques to give three-dimensional images of the beam, enabling beam analysis to be extended to laser beams with more complex distributions than purely gaussian.

Optilas, a distributor of lasers and related equipment, has introduced a new form of **laser safety spectacles** made by the German-based company LaserVision (Reader Service No. 111). The L-05 spectacle uses curved rather than flat lenses, lightweight glass with dielectric coatings to reflect the laser light rather than heavy, absorbing mineral glass lenses. The use of coated lenses is also said to give the spectacles a higher visible light transmission, so users of near-infrared diodes, titanium:sapphire and Nd:YAG lasers can have clear, colourless glasses for improved safety and comfort. □

These notes are compiled by Diane Gershon with information provided by the manufacturers. For more details on any of the products featured, fill in the reader service card bound inside the journal.